NON-PROVISIONAL PATENT APPLICATION

METHODS OF DETECTING SOFT TISSUE SARCOMA, COMPOSITIONS AND METHODS OF SCREENING FOR SOFT TISSUE SARCOMA MODULATORS

Inventor(s):

Natasha Aziz, a citizen of the United States, residing at 411 California Avenue; Palo Alto, California 94306;

Wendy M. Ginsburg, a citizen of the United States residing at 655 Page Street, San Francisco, California 94117; and

Albert Zlotnik, a citizen of the United States, residing at 507 Alger Drive, Palo Alto, California 94306

Assignee: Protein Design Labs, Inc.
34801 Campus Drive
Fremont, CA 94555

Entity: Large

HOWREY SIMON ARNOLD and WHITE, LLP
Box 34
301 Ravenswood Avenue
Menlo Park, California 94025-3453
Tel: 650-463-8100

METHODS OF DETECTING SOFT TISSUE SARCOMA, COMPOSITIONS AND METHODS OF SCREENING FOR SOFT TISSUE SARCOMA MODULATORS

PRIORITY INFORMATION

This application claims the benefit of U.S. Provisional Application No. 60/429,739, filed November 26, 2002.

FIELD OF THE INVENTION

The invention relates to the identification of nucleic acid and protein expression profiles and nucleic acids, products, and antibodies thereto that are involved in soft tissue sarcomas; and to the use of such expression profiles and compositions in diagnosis and therapy of such cancers. The invention further relates to methods for identifying and using agents and/or targets that modulate these cancers.

BACKGROUND OF THE INVENTION

Background on Soft Tissue Sarcomas is available, e.g., from Montgomery and Aaron (2001) Clinical Pathology of Soft-Tissue Tumors Marcel Dekker; ISBN: 0824702905; Brennan, et al. "Soft tissue sarcoma" pp 1738-1788 in DeVita, et al. (eds. 1997) Cancer: Principles and Practice of Oncology (5th ed.) Lippincott-Raven Philadelphia, PA; Pisters, et al. (2001) Cancer Management: A Multidisciplinary Approach (5th ed.) PRR; p 127-137 in American Joint Committee on Cancer (1992) Manual for Staging of Cancer (4th ed.) Lippincott, . Philadelphia; Schajowicz (1994) Tumors and Tumor-like Lesions of Bone: Pathology, Radiology and Treatment (2d ed.) Springer-Verlag, NY; Cotran, et al. (1999) Pathologic Basis of Disease Saunders; and various websites, e.g., NCI, Memorial Sloan-Kettering Cancer Center; cancerindex.com; cancersource.com; cancernetwork.com; and sarcoma.net.

Soft-tissue sarcomas are rare, representing only about 1 percent of all cancer cases. According to the American Cancer Society, approximately 8,700 new cases of soft-tissue sarcoma are diagnosed each year in adults and children in the United States. The age-adjusted incidence is 2 cases per 100,000 persons. There is a slight male predominance, with a male to female ratio of 11:10. The age distribution in

adult soft-tissue sarcoma studies is: <40 years, 20.7% of patients; 40-60 years, 27.6% of patients; > 60 years, 51.7% of patients.

In the United States, of the 8,700 new cases of soft-tissue sarcoma are identified annually, 4,400 patients die of the disease each year. The five-year survival percentages of soft tissue sarcomas range from 30% to 95% based on subtype and grade. The range for extremity sarcomas is 90%-95%, for trunk sarcomas 50%-75%, and for retroperitoneal lesions 30%-50%. In each of the three locations, higher-grade sarcomas have a poorer survival rate.

In a survey of approximately 5,000 soft-tissue sarcoma patients admitted to Memorial Sloan-Kettering Cancer Center from 1982 to 2001: 32 percent of sarcomas were found in the lower extremities; 18 percent in the viscera (organs located within the chest and abdomen, such as the stomach, kidney, uterus, etc.); 15 percent in the abdominal and retroperitoneal region; 13 percent in the upper extremities; 8 percent in the trunk and 14 percent in other sites.

The histological subtypes of soft tissue sarcomas include malignant fibrous histiocytoma, liposarcoma, fibrosarcoma, synovial sarcoma, rhabdomyosarcoma, and leiomyosarcoma. They occur over 50% of the time in extremities; the remainder occur in the head and neck and retroperitoneum. In addition, many of these tumors dedifferentiate. This results in a variety of overlapping patterns, making uniform classification difficult. The current histopathologic classification is based on the putative cell of origin of each lesion. Such classification based on histogenesis is reproducible for the more differentiated tumors. However, as the degree of histologic differentiation declines, it becomes increasingly difficult to determine cellular origin.

Liposarcomas are malignant tumors that develop from fat tissue. They can develop anywhere in the body, but they most often grow in the retroperitoneum (tissue at the back of the abdominal cavity). Fat tissue may also originate from other locations, usually in the arms, legs, or body cavities.

Rhabdomyosarcomas are malignant tumors that resemble developing skeletal muscle. These tumors most commonly grow in the arms or legs, but can also develop in the head or neck area, as well as the urinary and reproductive organs.

Synovial sarcomas are malignant tumors made up of cells that resemble the cells in joints. ("Synovial cells" line the joints.) However, synovial sarcomas do not necessarily arise in a joint, and the name is probably a misnomer, since the cancer

cells are probably quite different from normal joint cells. Synovial sarcomas can arise in any location in the body, and they often appear in young adults.

Fibrosarcomas are cancer of the fibroblast-type cells in the body. Fibroblasts form scars and do other important connective functions. Fibrosarcomas often occur in tendons and ligaments (fibrous tissue), usually in the arms, legs, or trunk. Fibrosarcomas are rare, accounting for fewer than 7% of primary malignant bone tumors. The five- and ten-year survival rates after radical surgery have been reported at 28% and 21.8%, respectively.

Chondrosarcomas are tumors of cells that form cartilage. Chondrosarcomas account for approximately 14% of malignant bone tumors. The incidence is greatest in individuals between 30 and 60 years of age, and among males. The most frequent sites of chondrosarcomas include the pelvic bone, long bones, scapula, and ribs. Less frequent sites include bones of the hand and foot, the nose, the maxilla, and the base of the skull. At present, chondrosarcomas remain nearly totally refractory to chemotherapeutic efforts inasmuch as chondrosarcomas usually have a poor blood supply. Consequently, drugs given intravenously generally do not reach the tumor in concentrations that are high enough to be therapeutically effective.

Malignant Fibrous Histiocytomas occur most commonly in the extremities (70-75%, with lower extremities accounting for 59% of cases), followed by the retroperitoneum. Tumors typically arise in deep fascia or skeletal muscle.

Leiomyosarcomas are malignant tumors that develop from smooth muscle tissue. They can arise anywhere in the body but the uterus or gastrointestinal tract are two relatively common locations.

In the majority of cases of soft-tissue sarcoma, no specific etiologic agent is identifiable. More commonly, an injury brings a preexisting neoplasm to the attention of the individual. However, a number of predisposing factors have been recognized.

Soft tissue sarcomas occur with greater frequency in patients with von Recklinghausen's disease (neurofibromatosis), Gardner's syndrome, Werner's syndrome, tuberous sclerosis, basal cell nevus syndrome, and among Li-Fraumeni kindreds (p53 mutations). The occurrences of bone tumors are also associated with hyperpara-thyroidism, chronic osteomyelitis, old bone infarct, osteochondromas, and enchondromas. Immunosuppressed patients such as renal transplant recipients and persons with autoimmunodeficiency syndrome (AIDS) have a higher risk for soft tissue sarcomas.

Soft-tissue sarcomas have been reported to originate in radiation fields following therapeutic radiation for a variety of solid tumors. Exposure to various chemicals in specific occupations or situations has been linked with the development of soft-tissue sarcoma. These include the phenoxy acetic acids (forestry and agriculture workers), chlorophenols (sawmill workers), Thorotrast (diagnostic x-ray technicians), vinyl chloride (individuals working with this gas, used in making plastics and as a refrigerant), and arsenic (vineyard workers).

Soft-tissue sarcomas have been reported after previous exposure to alkylating chemotherapeutic agents, most commonly after treatment of pediatric acute lymphocytic leukemia.

Related conditions include Reactive pseudosarcomatous proliferans (non-neoplastic lesions that mimic sarcomas), nodular fascilitis (infiltrative or pseudosarcomatous fascilitis), proliferative fascilitis, proliferative myositis, myositis ossificans, malignant giant cell tumor, malignant lymphoma of bone (reticulum cell sarcoma), Ewing's tumor (Ewing's sarcoma) and Osteosarcoma (osteogenic sarcoma).

Signs and symptoms of soft-tissue sarcoma depend, in large part, on the anatomic site of origin. Since 50% of soft-tissue sarcomas arise in an extremity, the majority of patients present with a palpable soft-tissue mass. Pain at presentation is noted in only one-third of cases.

Because there are so many varied subtypes, and because their characteristics are so different, the risk and seriousness of soft-tissue sarcomas can vary widely. In some patients, sarcomas are minor, non-threatening tumors that can be cured with simple surgical excision. In others, the tumors can be large and much more aggressive, and require chemotherapy and radiation therapy as well as surgery. In addition, the capacity of sarcomas to metastasize to other sites also varies widely. If metastasis occurs, it can sometimes be cured with surgery, but at other times it can be a life-threatening problem. In general, bone and soft tissue tumors tend to involve contiguous tissue and muscle, and aggressively metastasize early to the lungs via the hematogenous route. Occasionally, soft tissue sarcomas can spread to regional lymph nodes.

The prognosis for patients with adult soft tissue sarcomas depends on several factors, including the patient's age and the size, histologic grade, and stage of the tumor. Factors associated with a poorer prognosis include age older than 60 years of age, tumors larger than 5 centimeters, or high-grade histology. While low-grade

tumors are usually curable by surgery alone, higher-grade sarcomas (as determined by the mitotic index and the presence of hemorrhage and necrosis) are associated with higher local treatment failure rates and increased metastatic potential. Some histological subtypes such as rhabdomyosarcomas, synovial sarcomas, and malignant histiocytomas are considered poor prognosticators due to their high grade. If there is distant metastasis to the lymph nodes, lungs, or other bones, the prognosis is also lowered.

Standard treatment options include: Surgical excision, surgical excision with preoperative or postoperative radiation therapy, and if the tumor is unresectable, high-dose preoperative radiation therapy may be used, followed by surgical resection and postoperative radiation therapy. Today, doctors often give chemotherapy (Doxorubicin and ifosfamide) before surgery to patients with large, fast-growing sarcomas.

The development of advanced surgical techniques (e.g., microvascular tissue transfer, bone and joint replacement, and vascular reconstruction) and the application of multimodality approaches have allowed most patients to retain a functional extremity without any compromise in survival. Limb-sparing surgery employing adjuvant radiation to facilitate maximal local control has become the standard approach for large (T2) extremity soft-tissue sarcomas. In most centers, upwards of 90% of patients are treated with limb-sparing approaches. Amputation is reserved as a last resort option for local control, and is used with the knowledge that it does not affect survival.

Improved methods of diagnosis and prognosis of soft tissue sarcomas and effective treatment would be desirable. Accordingly, provided herein are methods that can be used in earlier diagnosis and prognosis of such cancers. Further provided are methods that can be used to screen candidate therapeutic agents for the ability to modulate, e.g., treat, them. Additionally provided herein are molecular targets and compositions for therapeutic intervention in these and other metastatic cancers.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods for detecting or modulating soft tissue sarcoma associated sequences.

In one aspect, the invention provides a method of detecting a sarcoma cancer-associated transcript in a cell in a patient, the method comprising contacting a

biological sample from the patient with a polynucleotide that selectively hybridized to a sequence at least 80% identical to a sequence as shown in Tables 1A-11C. In one embodiment, the biological sample is a tissue sample. In another embodiment, the biological sample comprises isolated nucleic acids, such as mRNA.

In another embodiment, the method further comprises the step of amplifying nucleic acids before the step of contacting the biological sample with the polynucleotide. Often, the polynucleotide comprises a sequence as shown in the Tables. The polynucleotide can be labeled, e.g., with a fluorescent label and can be immobilized on a solid surface.

In other embodiments the patient is undergoing a therapeutic regimen to treat a disease associated with these sarcomas or the patient is suspected of having a sarcoma-associated disorder.

In another aspect, the invention comprises an isolated nucleic acid molecule consisting of a polynucleotide sequence as shown in the Tables. The nucleic acid molecule can be labeled, e.g., with a fluorescent or radioactive label.

In other aspects, the invention provides an expression vector comprising an isolated nucleic acid molecule consisting of a polynucleotide sequence as shown in the Tables or a host cell comprising the expression vector.

In another aspect, the invention provides an isolated polypeptide which is encoded by a nucleic acid molecule having polynucleotide sequence as shown in Tables 1A-11C.

In another embodiment, the invention provides an antibody that specifically binds a polypeptide which is encoded by a nucleotide sequence of the Tables. The antibody can be conjugated or fused to an effector component such as a fluorescent label, a toxin, or a radioisotope. In some embodiments, the antibody is an antibody fragment or a humanized antibody.

In another aspect, the invention provides a method of detecting a cell undergoing such a cancer in a biological sample from a patient, the method comprising contacting the biological sample with an antibody that specifically binds to a polypeptide encoded by a nucleotide sequence of Tables 1A-11C. In some embodiments, the antibody is further conjugated or fused to an effector component, e.g., a fluorescent label.

In another embodiment, the invention provides a method of detecting antibodies specific to a sarcoma in a patient, the method comprising contacting a

biological sample from the patient with a polypeptide which is encoded by a nucleotide sequence of Tables 1A-11C.

The invention also provides a method of identifying a compound that modulates the activity of a sarcoma-associated polypeptide, the method comprising the steps of: (i) contacting the compound with a polypeptide encoded by a nucleotide sequence of Tables 1A-11C; and (ii) detecting an increase or a decrease in the activity of the polypeptide. In one embodiment, the polypeptide is encoded by a nucleotide sequence of Tables 1A-11C. In another embodiment, the polypeptide is expressed in a cell.

The invention also provides a method of identifying a compound that modulates the sarcoma, the method comprising steps of: (i) contacting the compound with a cell undergoing such cancer; and (ii) detecting an increase or a decrease in the expression of a polypeptide encoded by a nucleotide sequence of the Tables. In one embodiment, the detecting step comprises hybridizing a nucleic acid sample from the cell with a polynucleotide that selectively hybridizes to a sequence at least 80% identical to a sequence as shown in the Tables. In another embodiment, the method further comprises detecting an increase or decrease in the expression of a second sequence encoded by a nucleotide sequence of the Tables.

In another embodiment, the invention provides a method of inhibiting neoplastic properties in a cell that expresses a polypeptide at least 80% identical to a sequence encoded by a nucleotide sequence of Tables 1A-11C, the method comprising the step of contacting the cell with a therapeutically effective amount of an inhibitor of the polypeptide. In one embodiment, the polypeptide is encoded by a nucleotide sequence of Tables 1A-11C. In another embodiment, the inhibitor is an antibody.

Other aspects of the invention will become apparent by the following description of the invention.

Tables 1A-11C provide nucleotide sequence of genes that exhibit changes in expression levels as a function of time in tissue involved in cancer compared to normal or unaffected tissue.

DETAILED DESCRIPTION OF THE TABLES

Table 1A lists about 523 genes upregulated in chondrosarcoma relative to normal body tissues. These genes were selected from 59680 probesets on the

Eos/Affymetrix Hu03 GENECHIP® (DNA microchip) array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 2A lists about 763 genes upregulated in dermatofibrosarcoma protuberans relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 3A lists about 625 genes upregulated in fibrosarcoma relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 4A lists about 906 genes upregulated in liposarcoma relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 5A lists about 595 genes upregulated in synovial sarcoma relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 6A lists about 977 genes upregulated in rhabdomyosarcoma relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 7A lists about 973 genes upregulated in soft tissue sarcomas relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset

obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 8A lists about 712 genes upregulated in soft tissue sarcomas relative to normal soft tissues (muscle, skin, bone, adipose tissue). These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 9A lists about 1078 genes upregulated in malignant fibrous histiocytoma relative to normal body tissues. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression.

Table 10A lists about 501 genes upregulated in soft tissue sarcoma relative to normal body tissues that are likely to encode proteins amenable to modulation by small molecules, peptides, or antibodies. These genes were selected from 59680 probesets on the Eos/Affymetrix Hu03 GENECHIP® array. Gene expression data for each probeset obtained from this analysis was expressed as average intensity (AI), a normalized value reflecting the relative level of mRNA expression. The protein products of these genes often contain one or more domains indicative of have oncogenic function or of transducing intracellular signals, or of being modulatable by small molecules, peptides, or antibodies (e.g. pkinase, death-domain, 7tm, phosphatase, or ion_transporter). Certain predicted protein domains are noted.

Tables 1B-11B list the accession numbers for those Pkey's lacking UnigeneID's for tables 1A – 11A, respectively. For each probeset we have listed the gene cluster number from which the oligonucleotides were designed. Gene clusters were compiled using sequences derived from Genbank ESTs and mRNAs. These sequences were clustered based on sequence similarity using Clustering and Alignment Tools (DoubleTwist, Oakland California). The Genbank accession numbers for sequences comprising each cluster are listed in the "Accession" column.

Tables 1C-11C list the genomic positioning for those Pkey's lacking Unigene ID's and accession numbers in tables 1A-11A, respectively. For each predicted exon,

genomic sequence source used for prediction is also listed and cross-referenced. Nucleotide locations of each predicted exon are also listed.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

In accordance with the objects outlined above, the present invention provides novel methods for diagnosis and prognosis evaluation for soft tissue sarcomas cancer (sometimes referred to herein as sarcoma disorders or STSD), as well as methods for screening for compositions which modulate those cancers or similar disorders. Also provided are methods for treating these cancers and related conditions.

In particular, identification of markers selectively expressed on these cancers allows for use of that expression in diagnostic, prognostic, or therapeutic methods. As such, the invention defines various compositions, e.g., nucleic acids, polypeptides, antibodies, and small molecule agonists/antagonists, which will be useful to selectively recognize those markers. For example, therapeutic methods may take the form of protein therapeutics which use the marker expression for selective localization or modulation of function (for those markers which have a causative disease effect), for vaccines, identification of binding partners, or antagonism, e.g., using antisense or RNAi. The markers may be useful for molecular characterization of subsets of soft tissue sarcoma cancer or related diseases, which subsets may actually require very different treatments. Moreover, the markers may also be important in related diseases to the specific cancers, e.g., which affect similar tissues in non-malignant diseases, or have similar mechanisms of induction/maintenance. Metastatic processes or characteristics may also be targeted. Diagnostic and prognostic uses are made available, e.g., to subset related but distinct diseases, or to determine treatment strategy. The detection methods may be based upon nucleic acid, e.g., PCR or hybridization techniques, or protein, e.g., ELISA, imaging, IHC, etc. The diagnosis may be qualitative or quantitative, and may detect increases or decreases in expression levels.

Related conditions to these soft tissue sarcomas include, e.g., soft tissue tumors (e.g., fibrosarcoma, liposarcoma, leiomyosarcoma, histiocytoma, fibrohistiocytic sarcoma), smooth muscle tumors (e.g., rhabdomyoma, rhabdomyosarcoma) tumors of the blood and lymph vessels (e.g., angiosarcoma, lymphangiosarcoma, Kaposi's sarcoma), perivascular tumors (e.g., glomus tumors, hemangiopericytoma), synovial tumors (e.g., mesothelioma), neural tumors (e.g.,

neurofibroma, neurofibrosarcoma, malignant peripheral nerve sheath tumors, granular cell tumors, plexosarcoma, ganglioneuroblastoma, neuroepithelioma, extraskeletal Ewing's sarcoma, schwannoma, neuroma, ganglioneuroma), paraganglioma, extraskeletal cartilaginous and osseous tumors (e.g., chondrosarcoma, osteosarcoma), pluripotential mesenchymal tumors, epitheliod sarcomas, rhabdoid tumors, desmoplastic small cell tumors, and alveolar sarcomas. These markers may be similarly useful for addressing these related conditions, e.g., diagnosis, therapy, prognosis, etc.

Tables 1A-11C provide unigene cluster identification numbers for the nucleotide sequence of genes that exhibit increased or decreased expression in soft tissue sarcoma cancer samples. The tables also provide an exemplar accession number that provides a nucleotide sequence that is part of the unigene cluster.

Definitions

The term "soft tissue sarcoma cancer protein" or "soft tissue sarcoma cancer polynucleotide" refers to nucleic acid and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have a nucleotide sequence that has greater than about 60% nucleotide sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 92%, 94%, 96%, 97%, 98%, or 99% or greater nucleotide sequence identity, preferably over a region of over a region of at least about 25, 50, 100, 200, 500, 1000, or more nucleotides, to a nucleotide sequence of or associated with a unigene cluster of Tables 1A-11C; (2) bind to antibodies, e.g., polyclonal antibodies, raised against an immunogen comprising an amino acid sequence of Tables 1A-11C, and conservatively modified variants thereof; (3) specifically hybridize under stringent hybridization conditions to an anti-sense strand corresponding to a nucleic acid sequence of Tables 1A-11C and conservatively modified variants thereof; or (4) have an amino acid sequence that has greater than about 60% amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, preferably 90%, 91%, 93%, 95%, 97%, 98%, or 99% or greater amino sequence identity, preferably over a region of over a region of at least about 25, 50, 100, 200, 500, 1000, or more amino acids, to an amino acid sequence encoded by a nucleotide sequence of or associated with a unigene cluster of Tables 1A-11C. A polynucleotide or polypeptide sequence is typically from a mammal including, but not limited to, primate, e.g., human; rodent, e.g., rat, mouse, hamster; cow, pig, horse, sheep, or other mammal. A "soft tissue sarcoma cancer

polypeptide" and a "soft tissue sarcoma cancer polynucleotide," include both naturally occurring or recombinant forms.

A "full length" soft tissue sarcoma cancer protein or nucleic acid refers to a soft tissue sarcoma cancer polypeptide or polynucleotide sequence, or a variant thereof, that contains the elements normally contained in one or more naturally occurring, wild type soft tissue sarcoma cancer polynucleotide or polypeptide sequences. The "full length" may be prior to, or after, various stages of post-translation processing or splicing, including alternative splicing.

"Biological sample" as used herein is a sample of biological tissue or fluid that contains nucleic acids or polypeptides, e.g., of a cancer protein. Such samples include, but are not limited to, tissue isolated from primates, e.g., humans, or rodents, e.g., mice, and rats. Biological samples may also include sections of tissues such as biopsy and autopsy samples, and frozen sections taken for histologic purposes. A biological sample is typically obtained from a eukaryotic organism, most preferably a mammal such as a primate e.g., chimpanzee or human; cow; dog; cat; a rodent, e.g., guinea pig, rat, mouse; rabbit; or a bird; reptile; or fish. Livestock and domestic animals are of interest.

"Providing a biological sample" means to obtain a biological sample for use in methods described in this invention. Most often, this will be done by removing a sample of cells from a mammal, such as a human or animal, but can also be accomplished by using previously isolated cells (e.g., isolated by another person, at another time, and/or for another purpose), or by performing the methods of the invention in vivo. Archival tissues, having treatment or outcome history, will be particularly useful.

The terms "identical" or "percent identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (e.g., about 70% identity, preferably 75%, 80%, 85%, 90%, 91%, 93%, 95%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI web site or the like). Such sequences are then said to be "substantially identical." This definition also refers to,

or may be applied to, the compliment of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of one of the number of contiguous positions selected from the group consisting of from about 20-600, usually about 50-200, more usually about 100-150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482-489, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA 85:2444-2448, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel, et al. (eds. 1995 and supplements) Current Protocols in Molecular Biology Wiley).

Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include the BLAST and BLAST 2.0 algorithms, which are described in Altschul, et al. (1977) Nuc. Acids Res. 25:3389-3402; and Altschul, et al. (1990) J. Mol. Biol. 215:403-410. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the

nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul, et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA 89:10915-919) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences. See, e.g., Karlin and Altschul (1993) Proc. Nat'l Acad. Sci. USA 90:5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001. Log values may be negative large numbers, e.g., 5, 10, 20, 30, 40, 40, 70, 90, 110, 150, 170, etc.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, e.g., where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequences.

A "host cell" is a naturally occurring cell or a transformed cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be cultured cells, explants, cells in vivo, and the like. Host cells may be prokaryotic cells such as E. coli, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells, e.g., CHO, HeLa, and the like (see, e.g., the American Type Culture Collection catalog).

The terms "isolated," "purified," or "biologically pure" refer to material that is substantially or essentially free from components that normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein or nucleic acid that is the predominant species present in a preparation is substantially purified. In particular, an isolated nucleic acid is separated from some open reading frames that naturally flank the gene and encode proteins other than protein encoded by the gene. The term "purified" in some embodiments denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Preferably, it means that the nucleic acid or protein is at least about 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure. "Purify" or "purification" in other embodiments means removing at least one contaminant or component from the composition to be purified. In this sense, purification does not require that the purified compound be homogeneous, e.g., 100% pure.

The terms "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of

a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, e.g., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode a given protein. For instance, the codons GCA, GCC, GCG, and GCU all encode the amino acid alanine. Thus, at a position where an alanine is specified by a codon, the codon can be altered to one of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. In certain contexts each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only

codon for tryptophan) can be modified to yield a functionally similar molecule. Accordingly, a silent variation of a nucleic acid which encodes a polypeptide is implicit in a described sequence with respect to the expression product, but not necessarily with respect to actual probe sequences.

As to amino acid sequences, it will be recognized that individual substitutions, deletions, or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds, or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" particularly where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention. Typically conservative substitutions include for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M). See, e.g., Creighton (1984) Proteins: Structure and Molecular Properties Freeman).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, see, e.g., Alberts, et al. (eds. 2001) Molecular Biology of the Cell (4th ed.) Garland; and Cantor and Schimmel (1980) Biophysical Chemistry Part I: The Conformation of Biological Macromolecules Freeman. "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 25 to approximately 500 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed, usually by the noncovalent association of independent tertiary units. Anisotropic terms are also known as energy terms.

"Nucleic acid" or "oligonucleotide" or "polynucleotide" or grammatical equivalents used herein means at least two nucleotides covalently linked together. Oligonucleotides are typically from about 5, 6, 7, 8, 9, 10, 12, 15, 25, 30, 40, 50, or more nucleotides in length, up to about 100 nucleotides in length. Nucleic acids and polynucleotides are polymers, including longer lengths, e.g., 200, 300, 500, 1000, 2000, 3000, 5000, 7000, 10,000, etc. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs are included that may have at least one different linkage, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein (1992) Oligonucleotides and Analogues: A Practical Approach Oxford Univ. Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones, including those described in US Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7 of Sanghvi and Cook (eds. 1994) Carbohydrate Modifications in Antisense Research ACS Symposium Series 580. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

A variety of references disclose such nucleic acid analogs, including, e.g., phosphoramidate (Beaucage, et al. (1993) Tetrahedron 49:1925-1963 and references therein; Letsinger (1970) J. Org. Chem. 35:3800-3803; Sprinzl, et al. (1977) Eur. J. Biochem. 81:579-589; Letsinger, et al. (1986) Nuc. Acids Res. 14:3487-499; Sawai, et al. (1984) Chem. Lett. 805, Letsinger, et al. (1988) J. Am. Chem. Soc. 110:4470-4471; and Pauwels, et al. (1986) Chemica Scripta 26:141-149), phosphorothioate (Mag, et al. (1991) Nuc. Acids Res. 19:1437-441; and US Patent No. 5,644,048), phosphorodithioate (Brill, et al. (1989) J. Am. Chem. Soc. 111:2321-322), O-methylphophoroamidite linkages (see Eckstein (1992) Oligonucleotides and Analogues: A Practical Approach, Oxford Univ. Press), and peptide nucleic acid backbones and linkages (see Egholm (1992) J. Am. Chem. Soc. 114:1895-1897; Meier, et al. (1992) Chem. Int. Ed. Engl. 31:1008-1010; Nielsen (1993) Nature 365:566-568; and Carlsson, et al. (1996) Nature 380:207). Other analog nucleic acids

include those with positive backbones (Denpcy, et al. (1995) Proc. Nat'l Acad. Sci. USA 92:6097-101); non-ionic backbones (US Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141, and 4,469,863; Kiedrowski, et al. (1991) Angew. Chem. Intl. Ed. English 30:423-426; Letsinger, et al. (1988) J. Am. Chem. Soc. 110:4470-471; Jung, et al. (1994) Nucleoside and Nucleotide 13:1597-xxx; Chapters 2 and 3 in Sanghvi and Cook (eds. 1994) Carbohydrate Modifications in Antisense Research ACS Symposium Series 580; Mesmaeker, et al. (1994) Bioorganic and Medicinal Chem. Lett. 4:395-398; Jeffs, et al. (1994) J. Biomolecular NMR 34:17; and Horn, et al. (1996) Tetrahedron Lett. 37:743-xxx) and non-ribose backbones, including those described in US Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7 in Sanghvi and Cook (eds. 1994) Carbohydrate Modifications in Antisense Research ACS Symposium Series 580. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. See Jenkins, et al. (1995) Chem. Soc. Rev. pp 169-176. Several nucleic acid analogs are described in Rawls (page 35, June 2, 1997) C&E News.

Particularly useful are peptide nucleic acids (PNA) which includes peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in at least two advantages. The PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (T_m) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4° C drop in T_m for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to about 7-9° C. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration. In addition, PNAs are not degraded by cellular enzymes, and thus can be more stable.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The depiction of a single strand also defines the sequence of the complementary strand; thus the sequences described herein also provide the complement of the sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine,

xanthine hypoxanthine, isocytosine, isoguanine, etc. "Transcript" typically refers to a naturally occurring RNA, e.g., a pre-mRNA, hnRNA, or mRNA. As used herein, the term "nucleoside" includes nucleotides and nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus, e.g., the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, physiological, chemical, or other physical means. Useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, e.g., by incorporating a radiolabel into the peptide or used to detect antibodies specifically reactive with the peptide. The labels may be incorporated into the soft tissue sarcoma cancer nucleic acids, proteins, and antibodies. Many methods known for conjugating the antibody to the label may be employed. See, e.g., Hunter, et al. (1962) Nature 144:945; David, et al. (1974) Biochemistry 13:1014-1021; Pain, et al. (1981) J. Immunol. Meth. 40:219-230; and Nygren (1982) J. Histochem. and Cytochem. 30:407-412.

An "effector" or "effector moiety" or "effector component" is a molecule that is bound (or linked, or conjugated), either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds, to an antibody. The "effector" can be a variety of molecules including, e.g., detection moieties including radioactive compounds, fluorescent compounds, enzymes or substrates, tags such as epitope tags, toxins; activatable moieties, chemotherapeutic agents; chemoattractant or immunomodulating entities; lipases; antibiotics; or radioisotopes, e.g., emitting "hard" beta radiation.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, e.g., covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe. Alternatively, method using high affinity interactions may achieve the same results where one of a pair of binding partners binds to the other, e.g., biotin, streptavidin.

As used herein a "nucleic acid probe or oligonucleotide" is a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (e.g., A, C, G, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, preferably one that does not functionally interfere with hybridization. Thus, e.g., probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. Probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled, e.g., with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled, e.g., with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence. Diagnosis or prognosis may be based at the genomic level, or at the level of RNA or protein expression.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein, or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, e.g., recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed, or not expressed. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed in vitro, in general, by the manipulation of nucleic acid, e.g., using polymerases and endonucleases, in a form not normally found in nature. In this manner, operable linkage of different sequences is achieved. Thus an isolated nucleic acid, in a linear form, or an expression vector formed in vitro by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, e.g., using the in vivo cellular machinery of the host cell rather than in vitro manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Similarly, a "recombinant protein" is a protein made using recombinant techniques, e.g., through the expression of a recombinant nucleic acid as depicted above.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed in operable linkage to a promoter.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture

of nucleic acids, but to essentially no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in "Overview of principles of hybridization and the strategy of nucleic acid assays" in Tijssen (1993) Hybridization with Nucleic Probes (Laboratory Techniques in Biochemistry and Molecular Biology) (vol. 24) Elsevier. Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes (e.g., 10-50 nucleotides) and at least about 60° C for long probes (e.g., greater than about 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42° C, or, 5x SSC, 1% SDS, incubating at 65° C, with wash in 0.2x SSC, and 0.1% SDS at 65° C. For PCR, a temperature of about 36° C is typical for low stringency amplification, although annealing temperatures may vary between about 32-48° C depending on primer length. For high stringency PCR amplification, a temperature of about 62° C is typical, although high stringency annealing temperatures can range from about 50-65° C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90-95° C for 30-120 sec, an annealing phase lasting 30-120 sec, and an extension phase of about 72° C for 1-2 min. Protocols and guidelines for low and high stringency amplification reactions are provided, e.g., in Innis, et al. (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, NY.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37° C, and a wash in 1X SSC at 45° C. A positive hybridization is typically at least twice background. Alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency. Additional guidelines for determining hybridization parameters are provided in numerous references, e.g., Ausubel, et al. (eds. 1991 and supplements) Current Protocols in Molecular Biology Wiley.

The phrase "functional effects" in the context of assays for testing compounds that modulate activity of a soft tissue sarcoma cancer protein includes the determination of a parameter that is indirectly or directly under the influence of the cancer protein, e.g., a physiological, functional, physical, or chemical effect, such as the ability to increase or decrease soft tissue sarcoma cancer. It includes ligand binding activity; cell viability; cell growth on soft agar; anchorage dependence; contact inhibition and density limitation of growth; cellular proliferation; cellular transformation; growth factor or serum dependence; tumor specific marker levels; invasiveness into Matrigel; tumor growth and metastasis in vivo; mRNA and protein expression in cells undergoing metastasis; and other characteristics of cancer cells. "Functional effects" include in vitro, in vivo, and ex vivo activities.

By "determining the functional effect" is meant assaying for a compound that increases or decreases a parameter that is indirectly or directly under the influence of a soft tissue sarcoma cancer protein sequence, e.g., physiological, functional, enzymatic, physical, or chemical effects. Such functional effects can be measured, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties for the protein, measuring inducible markers or transcriptional activation of the cancer protein; measuring binding activity or binding assays, e.g., binding to antibodies or other ligands, and measuring growth, cellular proliferation, cell viability, cellular transformation, growth factor or serum dependence, tumor specific marker levels, invasiveness into Matrigel, tumor growth and metastasis in vivo, mRNA and

protein expression, and other characteristics of cancer cells. The functional effects can be evaluated by many means, e.g., microscopy for quantitative or qualitative measures of alterations in morphological features, measurement of RNA stability, identification of downstream or reporter gene expression (CAT, luciferase, β-gal, GFP and the like), e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, and ligand binding assays.

"Inhibitors", "activators", and "modulators" of soft tissue sarcoma cancer polynucleotide and polypeptide sequences are used to refer to activating, inhibitory, or modulating molecules identified using in vitro and in vivo assays of cancer polynucleotide and polypeptide sequences. Inhibitors are compounds that, e.g., bind to, partially or totally block activity, decrease, prevent, delay activation, inactivate, desensitize, or down regulate the activity or expression of soft tissue sarcoma cancer proteins, e.g., antagonists. "Activators" are compounds that increase, open, activate, facilitate, enhance activation, sensitize, agonize, or up regulate soft tissue sarcoma cancer protein activity. Inhibitors, activators, or modulators also include genetically modified versions of soft tissue sarcoma cancer proteins, e.g., versions with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, antibodies, small chemical molecules and the like. Such assays for inhibitors and activators include, e.g., expressing the cancer protein in vitro, in cells, or cell membranes, applying putative modulator compounds, and then determining the functional effects on activity, as described above. Activators and inhibitors of soft tissue sarcoma cancer can also be identified by incubating cancer cells with the test compound and determining increases or decreases in the expression of 1 or more soft tissue sarcoma cancer proteins, e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 or more cancer proteins, such as soft tissue sarcoma cancer proteins comprising the sequences set out in the Tables.

Samples or assays comprising soft tissue sarcoma cancer proteins that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of inhibition. Control samples (untreated with inhibitors) are assigned a relative protein activity value of 100%. Inhibition of a polypeptide is achieved when the activity value relative to the control is about 80%, preferably 50%, more preferably 25-0%. Activation of a soft tissue sarcoma cancer polypeptide is achieved when the activity value relative to the control (untreated with activators) is 110%, more preferably

150%, more preferably about 200-500% (e.g., two to five fold higher relative to the control), more preferably 1000-3000% higher.

The phrase "changes in cell growth" refers to a change in cell growth or proliferation characteristics in vitro or in vivo, such as cell viability, formation of foci, anchorage independence, semi-solid or soft agar growth, changes in contact inhibition and density limitation of growth, loss of growth factor or serum requirements, changes in cell morphology, gaining or losing immortalization, gaining or losing tumor specific markers, ability to form or suppress tumors when injected into suitable animal hosts, and/or immortalization of the cell. See, e.g., pp. 231-241 in Freshney (1994) Culture of Animal Cells a Manual of Basic Technique (2d ed.) Wiley-Liss.

"Tumor cell" refers to precancerous, cancerous, and normal cells in a tumor.

"Cancer cells," "transformed" cells or "transformation" in tissue culture, refers to spontaneous or induced phenotypic changes that do not necessarily involve the uptake of new genetic material. Although transformation can arise from infection with a transforming virus and incorporation of new genomic DNA, or uptake of exogenous DNA, it can also arise spontaneously or following exposure to a carcinogen, thereby mutating an endogenous gene. Transformation is associated with phenotypic changes, such as immortalization of cells, aberrant growth control, nonmorphological changes, and/or malignancy. See, Freshney (2001) Culture of Animal Cells: A Manual of Basic Technique (4th ed.) Wiley-Liss.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. Typically, the antigen-binding region of an antibody will be most critical in specificity and affinity of binding. See Paul (ed. 1999) Fundamental Immunology (4th ed.) Raven.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100-110 or more

amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, e.g., pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region. See Paul (ed. 1999) Fundamental Immunology (4th ed.) Raven. While various antibody fragments are defined in terms of the digestion of an intact antibody, it will be appreciated that such fragments may be synthesized de novo either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized de novo using recombinant DNA methodologies (e.g., single chain Fv) or those identified using phage display libraries (e.g., McCafferty, et al. (1990) Nature 348:552-554).

For preparation of antibodies, e.g., recombinant, monoclonal, or polyclonal antibodies, many techniques can be used. See, e.g., Kohler and Milstein (1975) Nature 256:495-497; Kozbor, et al. (1983) Immunology Today 4:72; Cole, et al. (1985) pp. 77-96 in Reisfeld and Sell (1985) Monoclonal Antibodies and Cancer Therapy Liss; Coligan (1991) Current Protocols in Immunology Lippincott; Harlow and Lane (1988) Antibodies: A Laboratory Manual CSH Press; and Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press. Techniques for the production of single chain antibodies (US Patent 4,946,778) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens. See, e.g., McCafferty, et al. (1990) Nature 348:552-554; Marks, et al. (1992) Biotechnology 10:779-783.

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, e.g., an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

Identification, Expression of soft tissue sarcoma cancer-associated sequences

In one aspect, the expression levels of genes are determined in different patient samples for which diagnosis information is desired, to provide expression profiles. An expression profile of a particular sample is essentially a "fingerprint" of the state of the sample; while two states may have a particular gene similarly expressed, the evaluation of a number of genes simultaneously allows the generation of a gene expression profile that is unique to the state of the cell. That is, normal tissue may be distinguished from sarcoma disorder tissue. By comparing expression profiles of tissue in known different soft tissue sarcoma cancer states, e.g., stages or disease course outcomes, information regarding which genes are important (including both up- and down-regulation of genes) in each of these states is obtained. Molecular profiling may distinguish subtypes of a currently collective disease designation, e.g., different forms of cancer processes.

The identification of sequences that are differentially expressed in cancer versus non-cancer tissue allows the use of this information in a number of ways. For example, a particular treatment regime may be evaluated: does a chemotherapeutic drug act to down-regulate soft tissue sarcoma cancer, and thus tumor growth or recurrence, in a particular patient. Alternatively, a treatment step may induce other markers which may be used as targets to destroy tumor cells. Similarly, diagnosis and treatment outcomes may be done or confirmed by comparing patient samples with the known expression profiles. Cancer tissue can be compared to non-cancerous conditions, or be analyzed to determine the stage of soft tissue sarcoma cancer in the tissue, or origin of primary tumor, e.g., metastasis from a remote primary site. Furthermore, these gene expression profiles (or individual genes) allow screening of drug candidates with an eye to mimicking or altering a particular expression profile; e.g., screening can be done for drugs that suppress the cancer expression profile. This

may be done by making biochips comprising sets of the important soft tissue sarcoma cancer genes, which can then be used in these screens. These methods can also be done on the protein basis; that is, protein expression levels of the cancer proteins can be evaluated for diagnostic purposes or to screen candidate agents. In addition, the cancer nucleic acid sequences can be administered for gene therapy purposes, including the administration of antisense nucleic acids, or the cancer proteins (including antibodies and other modulators thereof) administered as therapeutic drugs.

Thus the present invention provides nucleic acid and protein sequences that are differentially expressed in soft tissue sarcoma cancer relative to normal tissues and/or non-malignant disease, or in different types of related diseases, herein termed "soft tissue sarcoma cancer sequences". As outlined below, soft tissue sarcoma cancer sequences include those that are up-regulated (e.g., expressed at a higher level) in disorders associated with soft tissue sarcoma cancer, as well as those that are down-regulated (e.g., expressed at a lower level). In one embodiment, the soft tissue sarcoma cancer sequences are from humans; however, as will be appreciated, soft tissue sarcoma cancer sequences from other organisms may be useful in animal models of disease and drug evaluation; thus, other soft tissue sarcoma cancer sequences are provided, from vertebrates, including mammals, including rodents (rats, mice, hamsters, guinea pigs, etc.), primates, farm animals (including sheep, goats, pigs, cows, horses, etc) and pets (e.g., dogs, cats, etc.). Soft tissue sarcoma cancer sequences from other organisms may be obtained using the techniques outlined below.

Soft tissue sarcoma cancer sequences can include both nucleic acid and amino acid sequences. In one embodiment, the soft tissue sarcoma cancer sequences are recombinant nucleic acids. These nucleic acid sequences are useful in a variety of applications, including diagnostic applications, which will detect naturally occurring nucleic acids, as well as screening applications; e.g., biochips comprising nucleic acid probes or PCR microtiter plates with selected probes to the soft tissue sarcoma cancer sequences.

A soft tissue sarcoma cancer sequence can be initially identified by substantial nucleic acid and/or amino acid sequence homology to the sequences outlined herein. Such homology can be based upon the overall nucleic acid or amino acid sequence, and is generally determined as outlined below, e.g., using homology programs or hybridization conditions.

For identifying soft tissue sarcoma cancer-associated sequences, the cancer screen typically includes comparing genes identified in different tissues, e.g., normal and cancer tissues, cancer and non-malignant conditions, non-malignant conditions and normal tissues, or tumor tissue samples from patients who have metastatic disease vs. non metastatic tissue. Other suitable tissue comparisons include comparing cancer samples with metastatic cancer samples from other cancers, such as lung, stomach, gastrointestinal cancers, etc. Samples of different stages of cancer, e.g., survivor tissue, drug resistant states, and tissue undergoing metastasis, are applied to biochips comprising nucleic acid probes. The samples are first microdissected, if applicable, and treated for the preparation of mRNA. Suitable biochips are commercially available, e.g., from Affymetrix, Santa Clara, CA. Gene expression profiles as described herein are generated and the data analyzed.

In one embodiment, the genes showing changes in expression as between normal and disease states are compared to genes expressed in other normal tissues, including, and not limited to lung, heart, brain, liver, stomach, kidney, muscle, colon, small intestine, large intestine, spleen, bone, and/or placenta. In another embodiment, those genes identified during the cancer screen that are expressed in a significant amount in other tissues (e.g., essential organs) are removed from the profile, although in some embodiments, this is not necessary (e.g., where organs may be dispensable, e.g., female or male specific). That is, when screening for drugs, it is usually preferable that the target expression be disease specific, to minimize possible side effects on other organs were there expression.

In one embodiment, soft tissue sarcoma cancer sequences are those that are up-regulated in soft tissue sarcoma cancer; that is, the expression of these genes is higher in the cancer tissue as compared to non-cancer or non-malignant tissue. "Up-regulation" as used herein often means at least about a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being preferred. Another embodiment is directed to sequences up-regulated in non-malignant conditions relative to normal. Uniformity among relevant samples is desired.

Unigene cluster identification numbers and accession numbers herein are for the GenBank sequence database and the sequences of the accession numbers are hereby expressly incorporated by reference. GenBank is available, see, e.g., Benson, et al. (1998) Nuc. Acids Res. 26:1-7. Sequences are also available in other databases, e.g., European Molecular Biology Laboratory (EMBL) and DNA Database of Japan

(DDBJ). In some situations, the sequences may be derived from assembly of available sequences or be predicted from genomic DNA using exon prediction algorithms, such as FGENESH. See Salamov and Solovyev (2000) Genome Res. 10:516-522. In other situations, sequences have been derived from cloning and sequencing of isolated nucleic acids.

In another embodiment, soft tissue sarcoma cancer sequences are those that are down-regulated in cancer; that is, the expression of these genes is lower compared to non-cancer tissue. "Down-regulation" as used herein often means at least about a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being useful.

By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed in vitro, in general, by the manipulation of nucleic acid e.g., using polymerases and endonucleases, in a form not normally found in nature. Thus an isolated nucleic acid, in a linear form, or an expression vector formed in vitro by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, e.g., using the in vivo cellular machinery of the host cell rather than in vitro manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Similarly, a "recombinant protein" is a protein made using recombinant techniques, e.g., through the expression of a recombinant nucleic acid as depicted above. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. The protein may be isolated or purified away from some or most of the proteins and compounds with which it is normally associated in its wild type host, and thus may be substantially pure. An isolated protein is unaccompanied by at least some of the material with which it is normally associated in its natural state, preferably constituting at least about 0.5%, more preferably at least about 5% by weight of the total protein in a given sample. A substantially pure protein comprises at least about 75% by weight of the total protein, with at least about 80% being preferred, and at least about 90% being particularly preferred. The definition includes the production of a soft tissue sarcoma cancer protein from one organism in a different organism or host cell. Alternatively, the protein may be made

at a significantly higher concentration than is normally seen, through the use of an inducible promoter or high expression promoter, such that the protein is made at increased concentration levels. Alternatively, the protein may be in a form not normally found in nature, as in the addition of an epitope tag or amino acid substitutions, insertions and deletions, as discussed below.

In one embodiment, the soft tissue sarcoma cancer sequences are nucleic acids. Soft tissue sarcoma cancer sequences are useful in a variety of applications, including diagnostic applications, which will detect naturally occurring nucleic acids, as well as screening applications; e.g., biochips comprising nucleic acid probes to the cancer sequences can be generated. In the broadest sense, then, by "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs are included that may have alternate backbones, comprising, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein (1992) Oligonucleotides and Analogues: A Practical Approach, Oxford Univ. Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones, including those described in US Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7 in Sanghvi and Cook (eds. 1994) Carbohydrate Modifications in Antisense Research ACS Symposium Series 580. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip.

Nucleic acid analogs may find use in the present invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

Particularly useful are peptide nucleic acids (PNA) which includes peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting

temperature (T_m) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4° C drop in T_m for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9° C. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration. In addition, PNAs are not degraded by cellular enzymes, and thus can be more stable.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The depiction of a single strand also defines the sequence of a complementary strand; thus the sequences described herein also provide the complement of the sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc. As used herein, the term "nucleoside" includes nucleotides and nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus, e.g., the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

A soft tissue sarcoma cancer sequence can be initially identified by substantial nucleic acid and/or amino acid sequence homology to the soft tissue sarcoma cancer sequences outlined herein. Such homology can be based upon the overall nucleic acid or amino acid sequence, and is generally determined as outlined below, using either homology programs or hybridization conditions.

For identifying soft tissue sarcoma cancer-associated sequences, the cancer screen typically includes comparing genes identified in cancer cells with genes identified in controls. Samples of normal tissue and tissue associated with soft tissue sarcoma cancer are applied to biochips comprising nucleic acid probes. The samples are first microdissected, if applicable, and treated for the preparation of mRNA. Suitable biochips are commercially available, e.g., from Affymetrix. Gene expression profiles as described herein are generated and the data analyzed.

In one embodiment, the genes showing changes in expression as between normal and disease states are compared to genes expressed in other normal tissues, including, but not limited to lung, heart, brain, liver, breast, kidney, muscle, prostate,

small intestine, large intestine, spleen, bone, and placenta. In another embodiment, those genes identified during the cancer screen that are expressed in significant amount in other tissues are removed from the profile, although in some embodiments, this is not necessary. That is, when screening for drugs, it is usually preferable that the target be disease specific, to minimize possible side effects.

In one embodiment, soft tissue sarcoma cancer sequences are those that are up-regulated in soft tissue sarcoma cancer disorders; that is, the expression of these genes is higher in the disease tissue as compared to normal tissue. "Up-regulation" as used herein means at least about a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being preferred. Accession numbers herein are for the GenBank sequence database and the sequences of the accession numbers are hereby expressly incorporated by reference. See, e.g., Benson, et al. (1998) Nuc. Acids Res. 26:1-7. Sequences are also available in other databases, e.g., European Molecular Biology Laboratory (EMBL) and DNA Database of Japan (DDBJ). In addition, most genes were found to be expressed in a limited amount or not at all in heart, brain, lung, liver, breast, kidney, prostate, small intestine, and spleen.

In another embodiment, soft tissue sarcoma cancer sequences are those that are down-regulated in the soft tissue sarcoma cancer disorder; that is, the expression of these genes is lower in cancer tissue as compared to normal tissue. "Down-regulation" as used herein means at least about a two-fold change, preferably at least about a three fold change, with at least about five-fold or higher being preferred.

Informatics

The ability to identify genes that undergo changes in expression with time during soft tissue sarcoma cancer can additionally provide high-resolution, high-sensitivity datasets which can be used in the areas of diagnostics, therapeutics, drug development, biosensor development, and other related areas. For example, the expression profiles can be used in diagnostic or prognostic evaluation of patients with soft tissue sarcoma cancer-associated disease. Or as another example, subcellular toxicological information can be generated to better direct drug structure and activity correlation. See Anderson (June 11-12, 1998) Pharmaceutical Proteomics: Targets, Mechanism, and Function, paper presented at the IBC Proteomics conference, Coronado, CA. Subcellular toxicological information can also be utilized in a

biological sensor device to predict the likely toxicological effect of chemical exposures and likely tolerable exposure thresholds (see, US Patent No. 5,811,231). Similar advantages accrue from datasets relevant to other biomolecules and bioactive agents (e.g., nucleic acids, saccharides, lipids, drugs, and the like).

Thus, in another embodiment, the present invention provides a database that includes at least one set of data assay data. The data contained in the database is acquired , e.g., using array analysis either singly or in a library format. The database can be in a form in which data can be maintained and transmitted, but is preferably an electronic database. The electronic database of the invention can be maintained on an electronic device allowing for the storage of and access to the database, such as a personal computer, but is preferably distributed on a wide area network, such as the World Wide Web.

The focus of the present section on databases that include peptide sequence data is for clarity of illustration only. Similar databases can be assembled for assay data acquired using an assay of the invention.

The compositions and methods for identifying and/or quantitating the relative and/or absolute abundance of a variety of molecular and macromolecular species from a biological sample exhibiting soft tissue sarcoma cancer, e.g., the identification of soft tissue sarcoma cancer-associated sequences described herein, provide an abundance of information, which can be correlated with pathological conditions, predisposition to disease, drug testing, therapeutic monitoring, gene-disease causal linkages, identification of correlates of immunity and physiological status, among others. Although the data generated from the assays of the invention is suited for manual review and analysis, prior data processing using high-speed computers may be utilized.

An array of methods for indexing and retrieving biomolecular information is available. For example, US Patents 6,023,659 and 5,966,712 disclose a relational database system for storing biomolecular sequence information in a manner that allows sequences to be catalogued and searched according to one or more protein function hierarchies. US Patent 5,953,727 discloses a relational database having sequence records containing information in a format that allows a collection of partial-length DNA sequences to be catalogued and searched according to association with one or more sequencing projects for obtaining full-length sequences from the collection of partial length sequences. US Patent 5,706,498 discloses a gene database

retrieval system for making a retrieval of a gene sequence similar to a sequence data item in a gene database based on the degree of similarity between a key sequence and a target sequence. US Patent 5,538,897 discloses a method using mass spectroscopy fragmentation patterns of peptides to identify amino acid sequences in computer databases by comparison of predicted mass spectra with experimentally-derived mass spectra using a closeness-of-fit measure. US Patent 5,926,818 discloses a multi-dimensional database comprising a functionality for multi-dimensional data analysis described as on-line analytical processing (OLAP), which entails the consolidation of projected and actual data according to more than one consolidation path or dimension. US Patent 5,295,261 reports a hybrid database structure in which the fields of each database record are divided into two classes, navigational and informational data, with navigational fields stored in a hierarchical topological map which can be viewed as a tree structure or as the merger of two or more such tree structures. See also Mount (2001) Bioinformatics: Sequence and Genome Analysis CSH Press, NY; Durbin, et al. (eds. 1999) Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids Cambridge Univ. Press; Baxevanis and Oeullette (eds. 1998) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins (2d ed.) Wiley-Liss; Rashidi and Buehler (1999) Bioinformatics: Basic Applications in Biological Science and Medicine CRC Press; Setubal, et al. (eds. 1997) Introduction to Computational Molecular Biology Brooks/Cole; Misener and Krawetz (eds. 2000) Bioinformatics: Methods and Protocols Humana Press; Higgins and Taylor (eds. 2000) Bioinformatics: Sequence, Structure, and Databanks: A Practical Approach Oxford Univ. Press; Brown (2001) Bioinformatics: A Biologist's Guide to Biocomputing and the Internet Eaton Pub.; Han and Kamber (2000) Data Mining: Concepts and Techniques Kaufmann Pub.; and Waterman (1995) Introduction to Computational Biology: Maps, Sequences, and Genomes Chap and Hall.

The present invention provides a computer database comprising a computer and software for storing in computer-retrievable form assay data records cross-tabulated, e.g., with data specifying the source of the target-containing sample from which each sequence specificity record was obtained.

In an exemplary embodiment, at least one of the sources of target-containing sample is from a control tissue sample known to be free of pathological disorders. In a variation, at least one of the sources is a known pathological tissue specimen, e.g., a neoplastic lesion or another tissue specimen to be analyzed for soft tissue sarcoma

cancer. In another variation, the assay records cross-tabulate one or more of the following parameters for each target species in a sample: (1) a unique identification code, which can include, e.g., a target molecular structure and/or characteristic separation coordinate (e.g., electrophoretic coordinates); (2) sample source; and (3) absolute and/or relative quantity of the target species present in the sample.

The invention also provides for the storage and retrieval of a collection of target data in a computer data storage apparatus, which can include magnetic disks, optical disks, magneto-optical disks, DRAM, SRAM, SGRAM, SDRAM, RDRAM, DDR RAM, magnetic bubble memory devices, and other data storage devices, including CPU registers and on-CPU data storage arrays. Typically, the target data records are stored as a bit pattern in an array of magnetic domains on a magnetizable medium or as an array of charge states or transistor gate states, such as an array of cells in a DRAM device (e.g., each cell comprised of a transistor and a charge storage area, which may be on the transistor). In one embodiment, the invention provides such storage devices, and computer systems built therewith, comprising a bit pattern encoding a protein expression fingerprint record comprising unique identifiers for at least 10 target data records cross-tabulated with target source.

When the target is a peptide or nucleic acid, the invention preferably provides a method for identifying related peptide or nucleic acid sequences, comprising performing a computerized comparison between a peptide or nucleic acid sequence assay record stored in or retrieved from a computer storage device or database and at least one other sequence. The comparison can include a sequence analysis or comparison algorithm or computer program embodiment thereof (e.g., FASTA, TFASTA, GAP, BESTFIT) and/or the comparison may be of the relative amount of a peptide or nucleic acid sequence in a pool of sequences determined from a polypeptide or nucleic acid sample of a specimen.

The invention also may provide a magnetic disk, such as an IBM-compatible (DOS, Windows, Windows95/98/2000, Windows NT, OS/2) or other format (e.g., Linux, SunOS, Solaris, AIX, SCO Unix, VMS, MV, Macintosh, etc.) floppy diskette or hard (fixed, Winchester) disk drive, comprising a bit pattern encoding data from an assay of the invention in a file format suitable for retrieval and processing in a computerized sequence analysis, comparison, or relative quantitation method.

The invention also provides a network, comprising a plurality of computing devices linked via a data link, such as an Ethernet cable (coax or 10BaseT), telephone

line, ISDN line, wireless network, optical fiber, or other suitable signal transmission medium, whereby at least one network device (e.g., computer, disk array, etc.) comprises a pattern of magnetic domains (e.g., magnetic disk) and/or charge domains (e.g., an array of DRAM cells) composing a bit pattern encoding data acquired from an assay of the invention.

The invention also provides a method for transmitting assay data that includes generating an electronic signal on an electronic communications device, such as a modem, ISDN terminal adapter, DSL, cable modem, ATM switch, or the like, wherein the signal includes (in native or encrypted format) a bit pattern encoding data from an assay or a database comprising a plurality of assay results obtained by the method of the invention.

In one embodiment, the invention provides a computer system for comparing a query target to a database containing an array of data structures, such as an assay result obtained by the method of the invention, and ranking database targets based on the degree of identity and gap weight to the target data. A central processor is initialized to load and execute the computer program for alignment and/or comparison of the assay results. Data for a query target is entered into the central processor via an I/O device. Execution of the computer program results in the central processor retrieving the assay data from the data file, which comprises a binary description of an assay result.

The target data or record and the computer program can be transferred to secondary memory, which is typically random access memory (e.g., DRAM, SRAM, SGRAM, or SDRAM). Targets are ranked according to the degree of correspondence between a selected assay characteristic (e.g., binding to a selected affinity moiety) and the same characteristic of the query target and results are output via an I/O device. For example, a central processor can be a conventional computer (e.g., Intel Pentium, PowerPC, Alpha, PA-8000, SPARC, MIPS 4400, MIPS 10000, VAX, etc.); a program can be a commercial or public domain molecular biology software package (e.g., UWGCG Sequence Analysis Software, Darwin); a data file can be an optical or magnetic disk, a data server, a memory device (e.g., DRAM, SRAM, SGRAM, SDRAM, EPROM, bubble memory, flash memory, etc.); an I/O device can be a terminal comprising a video display and a keyboard, a modem, an ISDN terminal adapter, an Ethernet port, a punched card reader, a magnetic strip reader, or other suitable I/O device.

The invention also provides the use of a computer system, such as that described above, which comprises: (1) a computer; (2) a stored bit pattern encoding a collection of peptide sequence specificity records obtained by the methods of the invention, which may be stored in the computer; (3) a comparison target, such as a query target; and (4) a program for alignment and comparison, typically with rank-ordering of comparison results on the basis of computed similarity values.

Soft tissue sarcoma cancer-associated sequences

Soft tissue sarcoma cancer proteins of the present invention may be classified as secreted proteins, transmembrane proteins, or intracellular proteins. In one embodiment, the soft tissue sarcoma cancer protein is an intracellular protein. Intracellular proteins may be found in the cytoplasm and/or in the nucleus or associated with the intracellular side of the plasma membrane. Intracellular proteins are involved in all aspects of cellular function and replication (including, e.g., signaling pathways); aberrant expression of such proteins often results in unregulated or disregulated cellular processes. See, e.g., Alberts, et al. (eds. 1994) Molecular Biology of the Cell (3d ed.) Garland. For example, many intracellular proteins have enzymatic activity such as protein kinase activity, protein phosphatase activity, protease activity, nucleotide cyclase activity, polymerase activity, and the like. Intracellular proteins also serve as docking proteins that are involved in organizing complexes of proteins, or targeting proteins to various subcellular localizations, and are involved in maintaining the structural integrity of organelles.

An increasingly appreciated concept in characterizing proteins is the presence in the proteins of one or more motifs for which defined functions have been attributed. In addition to the highly conserved sequences found in the enzymatic domain of proteins, highly conserved sequences have been identified in proteins that are involved in protein-protein interaction. For example, Src-homology-2 (SH2) domains bind tyrosine-phosphorylated targets in a sequence dependent manner. PTB domains, which are distinct from SH2 domains, also bind tyrosine phosphorylated targets. SH3 domains bind to proline-rich targets. In addition, PH domains, tetratricopeptide repeats and WD domains to name only a few, have been shown to mediate protein-protein interactions. Some of these may also be involved in binding to phospholipids or other second messengers. These motifs can be identified on the basis of amino acid sequence; thus, an analysis of the sequence of proteins may

provide insight into both the enzymatic potential of the molecule and/or molecules with which the protein may associate. One useful database is Pfam (protein families), which is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Versions are available via the internet from Washington University in St. Louis, the Sanger Center in England, and the Karolinska Institute in Sweden. See, e.g., Bateman, et al. (2000) Nuc. Acids Res. 28:263-266; Sonnhammer, et al. (1997) Proteins 28:405-420; Bateman, et al. (1999) Nuc. Acids Res. 27:260-262; and Sonnhammer, et al. (1998) Nuc. Acids Res. 26:320-322.

In another embodiment, the cancer sequences are transmembrane proteins. Transmembrane proteins are molecules that span a phospholipid bilayer of a cell. They may have an intracellular domain, an extracellular domain, or both. The intracellular domains of such proteins may have a number of functions including those already described for intracellular proteins. For example, the intracellular domain may have enzymatic activity and/or may serve as a binding site for additional proteins. Frequently the intracellular domain of transmembrane proteins serves both roles. For example certain receptor tyrosine kinases have both protein kinase activity and SH2 domains. In addition, autophosphorylation of tyrosines on the receptor molecule itself, creates binding sites for additional SH2 domain containing proteins.

Transmembrane proteins may contain from one to many transmembrane domains. For example, receptor tyrosine kinases, certain cytokine receptors, receptor guanylyl cyclases and receptor serine/threonine protein kinases contain a single transmembrane domain. However, various other proteins including channels and adenylyl cyclases contain numerous transmembrane domains. Many important cell surface receptors such as G protein coupled receptors (GPCRs) are classified as "seven transmembrane domain" proteins, as they contain 7 membrane spanning regions. Characteristics of transmembrane domains include approximately 20 consecutive hydrophobic amino acids that may be followed or flanked by charged amino acids. Therefore, upon analysis of the amino acid sequence of a particular protein, the localization and number of transmembrane domains within the protein may be predicted. Important transmembrane protein receptors include, but are not limited to the insulin receptor, insulin-like growth factor receptor, human growth hormone receptor, glucose transporters, transferrin receptor, epidermal growth factor

receptor, low density lipoprotein receptor, epidermal growth factor receptor, leptin receptor, and interleukin receptors, e.g., IL-1 receptor, IL-2 receptor, etc.

The extracellular domains of transmembrane proteins are diverse; however, conserved motifs are found repeatedly among various extracellular domains. Conserved structure and/or functions have been ascribed to different extracellular motifs. Many extracellular domains are involved in binding to other molecules. In one aspect, extracellular domains are found on receptors. Factors that bind the receptor domain include circulating ligands, which may be peptides, proteins, or small molecules such as adenosine and the like. For example, growth factors such as EGF, FGF, and PDGF are circulating growth factors that bind to their cognate receptors to initiate a variety of cellular responses. Other factors include cytokines, mitogenic factors, neurotrophic factors and the like. Extracellular domains also bind to cell-associated molecules. In this respect, they mediate cell-cell interactions. Cell-associated ligands can be tethered to the cell, e.g., via a glycosylphosphatidylinositol (GPI) anchor, or may themselves be transmembrane proteins. Extracellular domains also associate with the extracellular matrix and contribute to the maintenance of the cell structure.

Soft tissue sarcoma cancer proteins that are transmembrane are useful in the present invention as they are readily accessible targets for immunotherapeutics, as are described herein. In addition, as outlined below, transmembrane proteins can be also useful in imaging modalities. Antibodies may be used to label such readily accessible proteins in situ. Alternatively, antibodies can also label intracellular proteins, in which case samples are typically permeabilized to provide access to intracellular proteins. In addition, some membrane proteins can be processed to release a soluble protein, or to expose a residual fragment. Released soluble proteins may be useful diagnostic markers, processed residual protein fragments may be useful markers of neoplastic disease.

A transmembrane protein can be made soluble by removing transmembrane sequences, e.g., through recombinant methods. Furthermore, transmembrane proteins that have been made soluble can be made to be secreted through recombinant means by adding an appropriate signal sequence.

In another embodiment, the cancer proteins are secreted proteins; the secretion of which can be either constitutive or regulated. These proteins have a signal peptide or signal sequence that targets the molecule to the secretory pathway. Secreted

proteins are involved in numerous physiological events; e.g., if circulating, they often serve to transmit signals to various other cell types. Secreted protein may function in an autocrine manner (acting on the cell that secreted the factor), a paracrine manner (acting on cells in close proximity to the cell that secreted the factor), an endocrine manner (acting on cells at a distance, e.g., secretion into the blood stream), or exocrine (secretion, e.g., through a duct or to adjacent epithelial surface as sweat glands, sebaceous glands, pancreatic ducts, lacrimal glands, mammary glands, wax producing glands of the ear, etc.). Thus secreted molecules often find use in modulating or altering numerous aspects of physiology. Soft tissue sarcoma cancer proteins that are secreted proteins are included in the present invention as they serve as good targets for diagnostic markers, e.g., for blood, plasma, serum, or stool tests. Those which are enzymes may be antibody or small molecule targets. Others may be useful as vaccine targets, e.g., via CTL mechanisms.

Soft tissue sarcoma cancer-associated nucleic acids

A soft tissue sarcoma cancer sequence is typically initially identified by substantial nucleic acid and/or amino acid sequence homology or linkage to the cancer sequences outlined herein. Such homology can be based upon the overall nucleic acid or amino acid sequence, and is generally determined as outlined below, using either homology programs or hybridization conditions. Typically, linked sequences on a mRNA are found on the same molecule.

As detailed elsewhere, percent identity can be determined using an algorithm such as BLAST. One method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively. Alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer nucleotides than those of the nucleic acids described, the percentage of homology may be determined based on the number of homologous nucleosides in relation to the total number of nucleosides. Thus, e.g., homology of sequences shorter than those of the sequences identified will be determined using the number of nucleosides in the shorter sequence.

In one embodiment, the nucleic acid homology is determined through hybridization studies. Thus, e.g., nucleic acids which hybridize under high stringency to a nucleic acid of Tables 1A-11C, or its complement, or is also found on naturally

occurring mRNAs is considered a soft tissue sarcoma cancer sequence. In another embodiment, less stringent hybridization conditions are used; e.g., moderate or low stringency conditions may be used; see Ausubel, supra, and Tijssen, supra.

The soft tissue sarcoma cancer nucleic acid sequences of the invention, e.g., the sequences in Tables 1A-11C, can be fragments of larger genes, e.g., they are nucleic acid segments. "Genes" in this context includes coding regions, non-coding regions, and mixtures of coding and non-coding regions. Accordingly, using the sequences provided herein, extended sequences, in either direction, of the cancer genes can be obtained, using techniques well known for cloning either longer sequences or the full length sequences; see Ausubel, et al., supra. Much can be done by informatics and many sequences can be clustered to include multiple sequences, e.g., systems such as UniGene.

Once the soft tissue sarcoma cancer nucleic acid is identified, it can be cloned and, if necessary, its constituent parts recombined to form the entire cancer nucleic acid coding regions or the entire mRNA sequence. Once isolated from its natural source, e.g., contained within a plasmid or other vector or excised therefrom as a linear nucleic acid segment, the recombinant cancer nucleic acid can be further-used as a probe to identify and isolate other soft tissue sarcoma cancer nucleic acids, e.g., extended coding regions. It can also be used as a "precursor" nucleic acid to make modified or variant cancer nucleic acids and proteins.

The soft tissue sarcoma cancer nucleic acids of the present invention are used in several ways. In one embodiment, nucleic acid probes to the cancer nucleic acids are made and attached to biochips to be used in screening and diagnostic methods, as outlined below, or for administration, e.g., for gene therapy, vaccine, and/or antisense applications. Alternatively, cancer nucleic acids that include coding regions of cancer proteins can be put into expression vectors for the expression of cancer proteins, again for screening purposes or for administration to a patient.

In another embodiment, nucleic acid probes to soft tissue sarcoma cancer nucleic acids (both the nucleic acid sequences outlined in the figures and/or the complements thereof) are made. The nucleic acid probes attached to the biochip are designed to be substantially complementary to cancer nucleic acids, e.g., the target sequence (either the target sequence of the sample or to other probe sequences, e.g., in sandwich assays), such that hybridization of the target sequence and the probes of the present invention occurs. As outlined below, this complementarity need not be

perfect; there may be a number of base pair mismatches which will interfere with hybridization between the target sequence and the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under normal reaction conditions, particularly high stringency conditions, as outlined herein.

A nucleic acid probe is generally single stranded but can be partially single and partially double stranded. The strandedness of the probe is dictated by the structure, composition, and properties of the target sequence. In general, the nucleic acid probes range from about 8-100 bases long, with from about 10-80 bases being preferred, and from about 30-50 bases being particularly preferred. That is, generally whole genes are not used. In some embodiments, much longer nucleic acids can be used, up to hundreds of bases.

In one embodiment, more than one probe per sequence is used, with either overlapping probes or probes to different sections of the target being used. That is, two, three, four, or more probes, with three being preferred, are used to build in a redundancy for a particular target. The probes can be overlapping (e.g., have some sequence in common), or separate. In some cases, PCR primers may be used to amplify signal for higher sensitivity.

Nucleic acids can be attached or immobilized to a solid support in a wide variety of ways. By "immobilized" and grammatical equivalents herein is meant the association or binding between the nucleic acid probe and the solid support is sufficient to be stable under the conditions of binding, washing, analysis, and removal as outlined. The binding can typically be covalent or non-covalent. By "non-covalent binding" and grammatical equivalents herein is meant one or more of electrostatic, hydrophilic, and hydrophobic interactions. Included in non-covalent binding is the covalent attachment of a molecule, such as, streptavidin to the support and the non-covalent binding of the biotinylated probe to the streptavidin. By "covalent binding" and grammatical equivalents herein is meant that the two moieties, the solid support and the probe, are attached by at least one bond, including sigma bonds, pi bonds and coordination bonds. Covalent bonds can be formed directly between the probe and the solid support or can be formed by a cross linker or by inclusion of a specific

reactive group on either the solid support or the probe or both molecules. Immobilization may also involve a combination of covalent and non-covalent interactions.

In general, the probes are attached to the biochip in a wide variety of ways. As described herein, the nucleic acids can either be synthesized first, with subsequent attachment to the biochip, or can be directly synthesized on the biochip.

The biochip comprises a suitable solid substrate. By "substrate" or "solid support" or other grammatical equivalents herein is meant a material that can be modified to contain discrete individual sites appropriate for the attachment or association of the nucleic acid probes and is amenable to at least one detection method. Often, the substrate may contain discrete individual sites appropriate for individual partitioning and identification. The number of possible substrates is very large, and includes, but is not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, TeflonJ, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, etc. In general, the substrates allow optical detection and do not appreciably fluoresce. See WO 00/55627.

Generally the substrate is planar, although other configurations of substrates may be used as well. For example, the probes may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Similarly, the substrate may be flexible, such as a flexible foam, including closed cell foams made of particular plastics.

In one embodiment, the surface of the biochip and the probe may be derivatized with chemical functional groups for subsequent attachment of the two. Thus, e.g., the biochip is derivatized with a chemical functional group including, but not limited to, amino groups, carboxy groups, oxo groups and thiol groups, with amino groups being particularly preferred. Using these functional groups, the probes can be attached using functional groups on the probes. For example, nucleic acids containing amino groups can be attached to surfaces comprising amino groups, e.g., using available linkers; e.g., homo-or hetero-bifunctional linkers as are well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages

155-200). In addition, in some cases, additional linkers, such as alkyl groups (including substituted and heteroalkyl groups) may be used.

In this embodiment, oligonucleotides are synthesized, and then attached to the surface of the solid support. Either the 5' or 3' terminus may be attached to the solid support, or attachment may be via an internal nucleoside. In another embodiment, the immobilization to the solid support may be very strong, yet non-covalent. For example, biotinylated oligonucleotides can be made, which bind to surfaces covalently coated with streptavidin, resulting in attachment.

Alternatively, the oligonucleotides may be synthesized on the surface. For example, photoactivation techniques utilizing photopolymerization compounds and techniques are used. In one embodiment, the nucleic acids can be synthesized in situ, using well known photolithographic techniques, such as those described in WO 95/25116; WO 95/35505; US Patent Nos. 5,700,637 and 5,445,934; and references cited within, all of which are expressly incorporated by reference; these methods of attachment form the basis of the Affymetrix GENECHIP® (DNA microchip array) technology.

Often, amplification-based assays are performed to measure the expression level of soft tissue sarcoma cancer-associated sequences. These assays are typically performed in conjunction with reverse transcription. In such assays, a soft tissue sarcoma cancer-associated nucleic acid sequence acts as a template in an amplification reaction (e.g., Polymerase Chain Reaction, or PCR). In a quantitative amplification, the amount of amplification product will be proportional to the amount of template in the original sample. Comparison to appropriate controls provides a measure of the amount of soft tissue sarcoma cancer-associated RNA. Methods of quantitative amplification are well known. Detailed protocols for quantitative PCR are provided, e.g., in Innis, et al. (1990) PCR Protocols: A Guide to Methods and Applications Academic Press.

In some embodiments, a TAQMAN® (reagents for nucleic acid amplification) based assay is used to measure expression. TAQMAN® based assays use a fluorogenic oligonucleotide probe that contains a 5' fluorescent dye and a 3' quenching agent. The probe hybridizes to a PCR product, but cannot itself be extended due to a blocking agent at the 3' end. When the PCR product is amplified in subsequent cycles, the 5' nuclease activity of the polymerase, e.g., AMPLITAQ®

(enzyme for diagnostic applications), results in the cleavage of the TAQMAN® probe. This cleavage separates the 5' fluorescent dye and the 3' quenching agent, thereby resulting in an increase in fluorescence as a function of amplification (see, e.g., literature provided by Perkin-Elmer).

Other suitable amplification methods include, but are not limited to, ligase chain reaction (LCR) (see Wu and Wallace (1989) Genomics 4:560-569, Landegren, et al. (1988) Science 241:1077-1080, and Barringer, et al. (1990) Gene 89:117-122), transcription amplification (Kwoh, et al. (1989) Proc. Nat'l Acad. Sci. USA 86:1173-1177), self-sustained sequence replication (Guatelli, et al. (1990) Proc. Nat'l Acad. Sci. USA 87:1874-1878), dot PCR, linker adapter PCR, etc.

Expression of soft tissue sarcoma cancer-associated proteins from nucleic acids

In one embodiment, soft tissue sarcoma cancer nucleic acids, e.g., encoding soft tissue sarcoma cancer proteins are used to make a variety of expression vectors to express cancer proteins which can then be used in screening assays, as described below. Expression vectors and recombinant DNA technology are well known (see, e.g., Ausubel, supra, and Fernandez and Hoeffler (eds. 1999) Gene Expression Systems Academic Press) and are used to express proteins. The expression vectors may be either self-replicating extrachromosomal vectors or vectors which integrate into a host genome. Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the soft tissue sarcoma cancer protein. The term "control sequences" refers to DNA sequences used for the expression of an operably linked coding sequence in a particular host organism. Control sequences that are suitable for prokaryotes, e.g., include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding

sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is typically accomplished by ligation at convenient restriction sites. If such sites do not exist, synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. Transcriptional and translational regulatory nucleic acid will generally be appropriate to the host cell used to express the soft tissue sarcoma cancer protein; e.g., transcriptional and translational regulatory nucleic acid sequences from Bacillus are preferably used to express the soft tissue sarcoma cancer protein in Bacillus. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known for a variety of host cells.

In general, transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In another embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known, and are useful in the present invention.

An expression vector may comprise additional elements. For example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, e.g., in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct. The integrating vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are available. See, e.g., Fernandez and Hoeffler, supra; and Kitamura, et al. (1995) Proc. Nat'l Acad. Sci. USA 92:9146-9150.

In addition, in another embodiment, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are available and will vary with the host cell used.

The soft tissue sarcoma cancer proteins of the present invention are produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a cancer protein, under the appropriate conditions to induce or cause expression of the cancer protein. Conditions appropriate for soft tissue sarcoma cancer protein expression will vary with the choice of the expression vector and the host cell, and will be easily ascertained through routine experimentation or optimization. For example, the use of constitutive promoters in the expression vector will require optimizing the growth and proliferation of the host cell, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important. For example, the baculoviral systems used in insect cell expression are lytic viruses, and thus harvest time selection can be crucial for product yield.

Appropriate host cells include yeast, bacteria, archaebacteria, fungi, and insect and animal cells, including mammalian cells. Of particular interest are Saccharomyces cerevisiae and other yeasts, E. coli, Bacillus subtilis, Sf9 cells, C129 cells, 293 cells, Neurospora, BHK, CHO, COS, HeLa cells, HUVEC (human umbilical vein endothelial cells), THP1 cells (a macrophage cell line), and various other human cells and cell lines.

In one embodiment, the soft tissue sarcoma cancer proteins are expressed in mammalian cells. Mammalian expression systems may be used, and include retroviral and adenoviral systems. One expression vector system is a retroviral vector system such as is generally described in PCT/US97/01019 and PCT/US97/01048. Of particular use as mammalian promoters are the promoters from mammalian viral genes, since the viral genes are often highly expressed and have a broad host range. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter, herpes simplex virus promoter, and the CMV promoter (see, e.g., Fernandez and Hoeffler, supra). Typically, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. Examples of transcription terminator and polyadenylation signals include those derived form SV40.

Methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, are available, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated

transfection, protoplast fusion, electroporation, viral infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

In one embodiment, soft tissue sarcoma cancer proteins are expressed in bacterial systems. Promoters from bacteriophage may also be used. In addition, synthetic promoters and hybrid promoters are also useful; e.g., the tac promoter is a hybrid of the trp and lac promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. In addition to a functioning promoter sequence, an efficient ribosome binding site is desirable. The expression vector may also include a signal peptide sequence that provides for secretion of the soft tissue sarcoma cancer protein in bacteria. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). The bacterial expression vector may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin, and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways. These components are assembled into expression vectors. Expression vectors for bacteria include vectors for Bacillus subtilis, E. coli, Streptococcus cremoris, and Streptococcus lividans, among others (e.g., Fernandez and Hoeffler, supra). The bacterial expression vectors are transformed into bacterial host cells using techniques such as calcium chloride treatment, electroporation, and others.

In one embodiment, soft tissue sarcoma cancer proteins are produced in insect cells using, e.g., expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors.

In another embodiment, a soft tissue sarcoma cancer protein is produced in yeast cells. Yeast expression systems include expression vectors for Saccharomyces cerevisiae, Candida albicans and C. maltosa, Hansenula polymorpha, Kluyveromyces fragilis and K. lactis, Pichia guillerimondii and P. pastoris, Schizosaccharomyces pombe, and Yarrowia lipolytica.

The soft tissue sarcoma cancer protein may also be made as a fusion protein, e.g., for the creation of monoclonal antibodies, if the desired epitope is small, the soft

tissue sarcoma cancer protein may be fused to a carrier protein to form an immunogen. Alternatively, the soft tissue sarcoma cancer protein may be made as a fusion protein to increase expression, or for other reasons. For example, when the soft tissue sarcoma cancer protein is a peptide, the nucleic acid encoding the peptide may be linked to another nucleic acid for expression purposes. Fusion with detection epitope tags can be made, e.g., with FLAG, His6, myc, HA, etc.

In one embodiment, the soft tissue sarcoma cancer protein is purified or isolated after expression. Soft tissue sarcoma cancer proteins may be isolated or purified in a variety of ways depending on what other components are present in the sample and the requirements for purified product. Standard purification methods include ammonium sulfate precipitations, electrophoretic, molecular, immunological, and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. For example, the cancer protein may be purified using a standard anti-cancer protein antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. See, e.g., Walsh (2002) Proteins: Biochemistry and Biotechnology Wiley; Hardin, et al. (eds. 2001) Cloning, Gene Expression and Protein Purification Oxford Univ. Press; Wilson, et al. (eds. 2000) Encyclopedia of Separation Science Academic Press; and Scopes (1993) Protein Purification Springer-Verlag. The degree of purification necessary will vary depending on the use of the cancer protein. In some instances no purification will be necessary.

Once expressed and purified if necessary, the soft tissue sarcoma cancer proteins and nucleic acids are useful in a number of applications. They may be used as immunoselection reagents, as vaccine reagents, as screening agents, etc.

In one embodiment, the soft tissue sarcoma cancer nucleic acids, proteins, and antibodies of the invention are labeled. By "labeled" herein is meant that a compound has at least one element, isotope, or chemical compound attached to enable the detection of the compound. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) immune labels, which may be antibodies, antigens, or epitope tags; and c) colored or fluorescent dyes. The labels may be incorporated into the cancer nucleic acids, proteins, and antibodies. For example, the label should be capable of producing, either directly or indirectly, a detectable signal. The detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C,

^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin, or an enzyme, such as alkaline phosphatase, beta-galactosidase, or horseradish peroxidase. Methods are known for conjugating the antibody to the label. See, e.g., Hunter, et al. (1962) Nature 144:945; David, et al. (1974) Biochemistry 13:1014-1021; Pain, et al. (1981) J. Immunol. Meth. 40:219-230; and Nygren (1982) J. Histochem. and Cytochem. 30:407-412.

Accordingly, the present invention also provides soft tissue sarcoma cancer protein sequences. A cancer protein of the present invention may be identified in several ways. "Protein" in this sense includes proteins, polypeptides, and peptides. Nucleic acid sequences of the invention can be used to generate protein sequences, including cloning the entire gene and verifying its frame and amino acid sequence, or by comparing it to known sequences to search for homology to provide a frame, assuming the soft tissue sarcoma cancer protein has an identifiable motif or homology to some protein in the database being used. Generally, the nucleic acid sequences are input into a program that will search all three frames for homology. This is done in an embodiment using the following NCBI Advanced BLAST parameters. The program is blastx or blastn. The database is nr. The input data is as "Sequence in FASTA format". The organism list is "none". The "expect" is 10; the filter is default. The "descriptions" is 500, the "alignments" is 500, and the "alignment view" is pairwise. The "Query Genetic Codes" is standard (1). The matrix is BLOSUM62; gap existence cost is 11, per residue gap cost is 1; and the lambda ratio is .85 default. This results in the generation of a putative protein sequence.

Variants of of soft tissue sarcoma cancer-associated proteins

Also included within one embodiment of soft tissue sarcoma cancer proteins are amino acid variants of the naturally occurring sequences, as determined herein. Preferably, the variants are preferably greater than about 75% homologous to the wild-type sequence, more preferably greater than about 80%, even more preferably greater than about 85%, and most preferably greater than 90%. In some embodiments the homology will be as high as about 93-95 or 98%. As for nucleic acids, homology in this context means sequence similarity or identity, with identity being preferred. This homology will be determined using standard techniques, as are outlined above for nucleic acid homologies.

Soft tissue sarcoma cancer proteins of the present invention may be shorter or longer than the wild type amino acid sequences. Thus, in one embodiment, included within the definition of soft tissue sarcoma cancer proteins are portions or fragments of the wild type sequences herein. In addition, as outlined above, the soft tissue sarcoma cancer nucleic acids of the invention may be used to obtain additional coding regions, and thus additional protein sequence.

In another embodiment, the soft tissue sarcoma cancer proteins are derivative or variant cancer proteins as compared to the wild-type sequence. That is, the derivative cancer peptide will often contain at least one amino acid substitution, deletion, or insertion, with amino acid substitutions being particularly useful at an appropriate position.

Also included within one embodiment of soft tissue sarcoma cancer proteins of the present invention are amino acid sequence variants. These variants typically fall into one or more of three classes: substitutional, insertional, or deletional variants. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding the cancer protein, using cassette or PCR mutagenesis or other appropriate techniques, to produce DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture as outlined above. However, variant cancer protein fragments having up to about 100-150 residues may be prepared by in vitro synthesis using established techniques. Amino acid sequence variants are characterized by the predetermined nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variation of the cancer protein amino acid sequence. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants can also be selected which have modified characteristics.

While the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed cancer variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, e.g., M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants is done using assays of cancer protein activities.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1-20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1-20 residues, although in some cases deletions may be much larger.

Substitutions, deletions, insertions or a combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the cancer protein are desired, substitutions are generally made in accordance with the amino acid substitution chart described.

The variants typically exhibit essentially the same qualitative biological activity and will elicit the same immune response as a naturally-occurring analog, although variants also are selected to modify the characteristics of cancer proteins as needed. Alternatively, the variant may be designed such that a biological activity of the cancer protein is altered. For example, glycosylation sites may be added, altered, or removed.

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those provided in the definition of "conservative substitution". For example, substitutions may be made which more significantly affect: the structure of the polypeptide backbone in the area of the alteration, e.g., the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic sidechain, e.g., serine or threonine, is substituted for (or by) a hydrophobic sidechain, e.g., leucine, isoleucine, phenylalanine, valine, or alanine; (b) a cysteine or proline is substituted for (or by) another residue; (c) a residue having an electropositive side chain, e.g., lysine, arginine, or histidine, is substituted for (or by) an electronegative side chain, e.g., glutamic or aspartic acid; (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine; or (e) a proline residue is incorporated or substituted, which changes the degree of rotational freedom of the peptidyl bond.

The variants typically exhibit a similar qualitative biological activity and will elicit the same immune response as the naturally-occurring analog, although variants

also are selected to modify the characteristics of the soft tissue sarcoma cancer proteins as needed. Alternatively, the variant may be designed such that the biological activity of the cancer protein is altered. For example, glycosylation sites may be altered or removed.

Covalent modifications of these cancer polypeptides are included within the scope of this invention. One type of covalent modification includes reacting targeted amino acid residues of a cancer polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N-or C-terminal residues of a cancer polypeptide. Derivatization with bifunctional agents is useful, e.g., for crosslinking cancer polypeptides to a water-insoluble support matrix or surface for use in the method for purifying anti-cancer polypeptide antibodies or screening assays. Commonly used crosslinking agents include, e.g., 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, e.g., esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutamine and asparagine residues to the corresponding glutamic and aspartic acid residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of serine, threonine, or tyrosine residues, methylation of the γ-amino groups of lysine, arginine, and histidine side chains (e.g., pp. 79-86, Creighton (1992) Proteins: Structure and Molecular Properties Freeman), acetylation of the N-terminal amine, and amidation of a C-terminal carboxyl group.

Another type of covalent modification of a cancer polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence cancer polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence cancer polypeptide. Glycosylation patterns can be altered in many ways. Different cell types may be used to express cancer-associated sequences to exhibit different glycosylation patterns.

Addition of glycosylation sites to soft tissue sarcoma cancer polypeptides may also be accomplished by altering the amino acid sequence thereof. The alteration may be made, e.g., by the addition of, or substitution by, one or more serine or threonine residues to the native sequence cancer polypeptide (for O-linked glycosylation sites). The soft tissue sarcoma cancer amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the cancer polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the soft tissue sarcoma cancer polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. See, e.g., WO 87/05330; and pp. 259-306 in Aplin and Wriston (1981) CRC Crit. Rev. Biochem.

Removal of carbohydrate moieties present on the soft tissue sarcoma cancer polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are applicable. See, e.g., Sojar and Bahl (1987) Arch. Biochem. Biophys. 259:52-57; and Edge, et al. (1981) Anal. Biochem. 118:131-137. Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo- and exo-glycosidases. See, e.g., Thotakura, et al. (1987) Meth. Enzymol. 138:350-359.

Another type of covalent modification of soft tissue sarcoma cancer protein comprises linking the cancer polypeptide to one of a variety of nonproteinaceous polymers, e.g., polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in US Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192; or 4,179,337.

Soft tissue sarcoma cancer polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a cancer polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a cancer polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxyl-terminus of the soft tissue sarcoma cancer polypeptide. The presence of such epitope-tagged forms of a cancer polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the soft tissue sarcoma cancer polypeptide

to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a cancer polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are available. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; HIS6, and metal chelation tags, the flu HA tag polypeptide and its antibody 12CA5 (Field, et al. (1988) Mol. Cell. Biol. 8:2159-2165); the c-myc tag and the 8F9, 3C7, 6E10, G4, B7, and 9E10 antibodies thereto (Evan, et al. (1985) Mol. Cell. Biol. 5:3610-3616); and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody (Paborsky, et al. (1990) Protein Engineering 3:547-553). Other tag polypeptides include the Flag-peptide (Hopp, et al. (1988) BioTechnology 6:1204-1210); the KT3 epitope peptide (Martin, et al. (1992) Science 255:192-194); tubulin epitope peptide (Skinner, et al. (1991) J. Biol. Chem. 266:15163-15166); and the T7 gene 10 protein peptide tag (Lutz-Freyermuth, et al. (1990) Proc. Nat'l Acad. Sci. USA 87:6393-6397).

Also included with an embodiment of soft tissue sarcoma cancer protein are other soft tissue sarcoma cancer proteins of the functional family, and counterpart cancer proteins from other organisms, which are cloned and expressed as outlined below. Thus, probe or degenerate polymerase chain reaction (PCR) primer sequences may be used to find other related cancer proteins from humans or other organisms. Particularly useful probe and/or PCR primer sequences include unique areas of the soft tissue sarcoma cancer nucleic acid sequence. PCR primers are from about 15-35 nucleotides in length, with from about 20-30 being preferred, and may contain inosine as needed. The conditions for the PCR reaction are well known. See, e.g., Innis, PCR Protocols, supra.

In addition, as is outlined herein, soft tissue sarcoma cancer proteins can be made that are longer than those encoded by the nucleic acids of the Tables, e.g., by the elucidation of extended sequences, the addition of epitope or purification tags, the addition of other fusion sequences, etc.

Soft tissue sarcoma cancer proteins may also be identified as being encoded by soft tissue sarcoma cancer nucleic acids. Thus, soft tissue sarcoma cancer proteins are

encoded by nucleic acids that will hybridize to the sequences of the sequence listings, or their complements, as outlined herein.

Antibodies to soft tissue sarcoma cancer-associated proteins

In one embodiment, when the soft tissue sarcoma cancer protein is to be used to generate antibodies, e.g., for immunotherapy or immunodiagnosis, the soft tissue sarcoma cancer protein should share at least one epitope or determinant with the full length protein. By "epitope" or "determinant" herein is typically meant a portion of a protein which will generate and/or bind an antibody or T-cell receptor in the context of MHC. Thus, in most instances, antibodies made to a smaller, e.g., fragment of, cancer protein will be able to bind to the full-length protein, particularly linear epitopes. In one embodiment, the epitope is unique; that is, antibodies generated to a unique epitope show little or no cross-reactivity. In another embodiment, the epitope is selected from a protein sequence set out in Tables 1A-11C.

Methods of preparing polyclonal antibodies exist (e.g., Coligan, supra; and Harlow and Lane, supra). Polyclonal antibodies can be raised in a mammal, e.g., by one or more injections of an immunizing agent and, if desired, an adjuvant. Typically, the immunizing agent and/or adjuvant will be injected in the mammal by multiple subcutaneous or intraperitoneal injections. The immunizing agent may include a protein encoded by a nucleic acid of the figures or fragment thereof or a fusion protein thereof. It may be useful to conjugate the immunizing agent to a protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. Examples of adjuvants which may be employed include Freund's complete adjuvant and MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate). The immunization protocol may be selected without undue experimentation.

The antibodies may, alternatively, be monoclonal antibodies. Monoclonal antibodies may be prepared using hybridoma methods, such as those described by Kohler and Milstein (1975) Nature 256:495-497. In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized in vitro. The immunizing agent will typically include a polypeptide

encoded by a nucleic acid of Tables 1A-11C, or fragment thereof, or a fusion protein thereof. Generally, either peripheral blood lymphocytes ("PBLs") are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (e.g., pp. 59-103 in Goding (1986) Monoclonal Antibodies: Principles and Practice Academic Press). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells may be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

In one embodiment, the antibodies are bispecific antibodies. Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens or that have binding specificities for two epitopes on the same antigen. In one embodiment, one of the binding specificities is for a protein encoded by a nucleic acid Tables 1A-11C or a fragment thereof, the other one is for another antigen, and preferably for a cell-surface protein or receptor or receptor subunit, preferably one that is tumor specific. Alternatively, tetramer-type technology may create multivalent reagents.

In one embodiment, the antibodies to soft tissue sarcoma cancer protein are capable of reducing or eliminating a biological function of a soft tissue sarcoma cancer protein, as is described below. That is, the addition of anti-soft tissue sarcoma cancer protein antibodies (either polyclonal or preferably monoclonal) to cancer tissue may reduce or eliminate the neoplastic or malignant cancer activity. Generally, at least about 25% decrease in activity, growth, size or the like may be used, with at least about 50% being particularly useful and about 95-100% decrease being especially useful.

In another embodiment the antibodies to the soft tissue sarcoma cancer proteins are humanized antibodies (e.g., Xenerex Biosciences; Medarex, Inc.; Abgenix, Inc.; Protein Design Labs, Inc.). Humanized forms of non-human (e.g.,

murine) antibodies are chimeric molecules of immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues form a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, a humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework (FR) regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will typically comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones, et al. (1986) Nature 321:522-525; Riechmann, et al. (1988) Nature 332:323-329; and Presta (1992) Curr. Op. Struct. Biol. 2:593-596). Humanization can be essentially performed following the method of Winter and co-workers (Jones, et al. (1986) Nature 321:522-525; Riechmann, et al. (1988) Nature 332:323-327; Verhoeyen, et al. (1988) Science 239:1534-1536), by substituting rodent CDRs or CDR sequences for corresponding sequences of a human antibody. Accordingly, such humanized antibodies are chimeric antibodies (US Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by corresponding sequence from a non-human species.

Human-like antibodies can also be produced using phage display libraries (Hoogenboom and Winter (1992) J. Mol. Biol. 227:381-388; Marks, et al. (1991) J. Mol. Biol. 222:581-597) or human monoclonal antibodies (e.g., p. 77, Cole, et al. in Reisfeld and Sell (1985) Monoclonal Antibodies and Cancer Therapy Liss; and Boerner, et al. (1991) J. Immunol. 147:86-95). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely

inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in nearly all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, e.g., in US Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks, et al. (1992) Bio/Technology 10:779-783; Lonberg, et al. (1994) Nature 368:856-859; Morrison (1994) Nature 368:812-13; Fishwild, et al. (1996) Nature Biotechnology 14:845-851, commented on in Neuberger (1996) Nature Biotechnology 14:826; and Lonberg and Huszar (1995) Intern. Rev. Immunol. 13:65-93.

By immunotherapy is meant treatment of soft tissue sarcoma cancer or related disease with an antibody raised against, e.g., proteins. As used herein, immunotherapy can be passive or active. Passive immunotherapy as defined herein is the passive transfer of antibody to a recipient (patient). Active immunization is the induction of antibody and/or T-cell responses in a recipient (patient). Induction of an immune response is the result of providing the recipient with an antigen to which antibodies are raised. The antigen may be provided by injecting a polypeptide against which antibodies are desired to be raised into a recipient, or contacting the recipient with a nucleic acid capable of expressing the antigen and under conditions for expression of the antigen, leading to an immune response.

In one embodiment the soft tissue sarcoma cancer proteins against which antibodies are raised are secreted proteins as described above. Without being bound by theory, antibodies used for treatment may bind and prevent the secreted protein from binding to its receptor, thereby inactivating the secreted cancer protein, e.g., in autocrine signaling.

In another embodiment, the soft tissue sarcoma cancer protein to which antibodies are raised is a transmembrane protein. Without being bound by theory, antibodies used for treatment often bind the extracellular domain of the cancer protein and prevent it from binding to other proteins, such as circulating ligands or cell-associated molecules. The antibody may cause down-regulation of the transmembrane cancer protein. The antibody may be a competitive, non-competitive, or uncompetitive inhibitor of protein binding to the extracellular domain of the cancer protein. The antibody may be an antagonist of the cancer protein. Further, the antibody prevents activation of the transmembrane cancer protein. In one aspect, when the antibody prevents the binding of other molecules to the cancer protein, the

antibody prevents growth of the cell. The antibody may also be used to target or sensitize the cell to cytotoxic agents, including, but not limited to TNF-α, TNF-β, IL-1, INF-γ, and IL-2, or chemotherapeutic agents including 5FU, vinblastine, actinomycin D, cisplatin, methotrexate, and the like. In some instances the antibody belongs to a sub-type that activates serum complement, or a similar effector function, when complexed with the transmembrane protein thereby mediating cytotoxicity or antigen-dependent cytotoxicity (ADCC). Thus, soft tissue sarcoma cancer is treated by administering to a patient antibodies directed against the transmembrane cancer protein. Antibody-labeling may activate a co-toxin, localize a toxin payload, or otherwise provide means to locally ablate cells.

In another embodiment, the antibody is conjugated or fused to an effector moiety. The effector moiety can be a labeling moiety, e.g., a radioactive or fluorescent label, or a therapeutic moiety. In one aspect the therapeutic moiety is a small molecule that modulates the activity of the soft tissue sarcoma cancer protein. In another aspect the therapeutic moiety modulates the activity of molecules associated with or in close proximity to the soft tissue sarcoma cancer protein. The therapeutic moiety may inhibit enzymatic activity such as protease or collagenase activity associated with the cancer, or be an attractant of other cells, such as NK cells. See, e.g., Groh, et al. (2002) Nature 419:734-738.

In one embodiment, the therapeutic moiety can also be a cytotoxic agent. In this method, targeting the cytotoxic agent to soft tissue sarcoma cancer tissue or cells, results in a reduction in the number of afflicted cells, thereby reducing symptoms associated with cancer. Cytotoxic agents are numerous and varied and include, but are not limited to, cytotoxic drugs or toxins or active fragments of such toxins. Suitable toxins and their corresponding fragments include diphtheria A chain, exotoxin A chain, ricin A chain, abrin A chain, curcin, crotin, phenomycin, enomycin, and the like. Cytotoxic agents also include radiochemicals made by conjugating radioisotopes to antibodies raised against soft tissue sarcoma cancer proteins, or binding of a radionuclide to a chelating agent that has been covalently attached to the antibody. Targeting the therapeutic moiety to transmembrane cancer proteins not only serves to increase the local concentration of therapeutic moiety in the cancer afflicted area, but also serves to reduce deleterious side effects that may be associated with the therapeutic moiety.

In another embodiment, the soft tissue sarcoma cancer protein against which the antibodies are raised is an intracellular protein. In this case, the antibody may be conjugated or fused to a protein which facilitates entry into the cell. In one case, the antibody enters the cell by endocytosis. In another embodiment, a nucleic acid encoding the antibody is administered to the individual or cell. Moreover, wherein the soft tissue sarcoma cancer protein can be targeted within a cell, e.g., the nucleus, an antibody thereto contains a signal for that target localization, e.g., a nuclear localization signal.

The soft tissue sarcoma cancer antibodies of the invention specifically bind to soft tissue sarcoma cancer proteins. By "specifically bind" herein is meant that the antibodies bind to the protein with a K_d of at least about 0.1 mM, more usually at least about 1 μM, preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better. Selectivity of binding to the specific target and not to related sequences is often also important

Detection of soft tissue sarcoma cancer-associated sequence for diagnostic and therapeutic applications

In one aspect, the RNA expression levels of genes are determined for different cellular states in the soft tissue sarcoma cancer phenotype. Expression levels of genes in normal tissue (e.g., not exhibiting soft tissue sarcoma cancer) and in soft tissue sarcoma cancer tissue (and in some cases, for varying severities of soft tissue sarcoma cancer that relate to prognosis, as outlined below) are evaluated to provide expression profiles. An expression profile of a particular cell state or point of development is essentially a "fingerprint" of the state. While two states may have a particular gene similarly expressed, the evaluation of a number of genes simultaneously allows the generation of a gene expression profile that is reflective of the state of the cell. By comparing expression profiles of cells in different states, information regarding which genes are important (including both up- and down-regulation of genes) in each of these states is obtained. Then, diagnosis may be performed or confirmed to determine whether a tissue sample has the gene expression profile of normal or cancer tissue. This will provide for molecular diagnosis of related conditions.

"Differential expression," or grammatical equivalents as used herein, refers to qualitative or quantitative differences in the temporal and/or cellular gene expression

patterns within and among cells and tissue. Thus, a differentially expressed gene can qualitatively have its expression altered, including an activation or inactivation, in, e.g., normal versus cancer tissue. Genes may be turned on or turned off in a particular state, relative to another state thus permitting comparison of two or more states. A qualitatively regulated gene will exhibit an expression pattern within a state or cell type which is detectable by standard techniques. Some genes will be expressed in one state or cell type, but not in both. Alternatively, the difference in expression may be quantitative, e.g., in that expression is increased or decreased; e.g., gene expression is either upregulated, resulting in an increased amount of transcript, or downregulated, resulting in a decreased amount of transcript. The degree to which expression differs need only be large enough to quantify via standard characterization techniques, e.g., as by use of Affymetrix GENECHIP® expression arrays. See, Lockhart (1996) Nature Biotechnology 14:1675-1680. Other techniques include, but are not limited to, quantitative reverse transcriptase PCR, Northern analysis and RNase protection. As outlined above, preferably the change in expression (e.g., upregulation or downregulation) is at least about 50%, more preferably at least about 100%, more preferably at least about 150%, more preferably at least about 200%, with from 300 to at least 1000% being especially useful.

Evaluation may be at the gene transcript, or the protein level. The amount of gene expression may be monitored using nucleic acid probes to the DNA or RNA equivalent of the gene transcript, and the quantification of gene expression levels, or, alternatively, the final gene product itself (protein) can be monitored, e.g., with antibodies to the soft tissue sarcoma cancer protein and standard immunoassays (ELISAs, etc.) or other techniques, including mass spectroscopy assays, 2D gel electrophoresis assays, etc. Proteins corresponding to soft tissue sarcoma cancer genes, e.g., those identified as being important in a soft tissue sarcoma cancer phenotype, can be evaluated in a soft tissue sarcoma cancer diagnostic test. In one embodiment, gene expression monitoring is performed simultaneously on a number of genes. Multiple protein expression monitoring can be performed as well. Similarly, these assays may be performed on an individual basis as well.

In this embodiment, the soft tissue sarcoma cancer nucleic acid probes are attached to biochips as outlined herein for the detection and quantification of soft

tissue sarcoma cancer sequences in a particular cell. The assays are further described below in the example. PCR techniques can be used to provide greater sensitivity.

In one embodiment nucleic acids encoding the soft tissue sarcoma cancer protein are detected. Although DNA or RNA encoding the soft tissue sarcoma cancer protein may be detected, of particular interest are methods wherein an mRNA encoding a cancer protein is detected. Probes to detect mRNA can be a nucleotide/deoxynucleotide probe that is complementary to and hybridizes with the mRNA and includes, but is not limited to, oligonucleotides, cDNA or RNA. Probes also should contain a detectable label, as defined herein. In one method the mRNA is detected after immobilizing the nucleic acid to be examined on a solid support such as nylon membranes and hybridizing the probe with the sample. Following washing to remove the non-specifically bound probe, the label is detected. In another method detection of the mRNA is performed in situ. In this method permeabilized cells or tissue samples are contacted with a detectably labeled nucleic acid probe for sufficient time to allow the probe to hybridize with the target mRNA. Following washing to remove the non-specifically bound probe, the label is detected. For example, a digoxygenin labeled riboprobe (RNA probe) that is complementary to the mRNA encoding a cancer protein is detected by binding the digoxygenin with an anti-digoxygenin secondary antibody and developed with nitro blue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate.

In one embodiment, various proteins from the three classes of proteins as described herein (secreted, transmembrane or intracellular proteins) are used in diagnostic assays. The soft tissue sarcoma cancer proteins, antibodies, nucleic acids, modified proteins and cells containing cancer sequences are used in diagnostic assays. This can be performed on an individual gene or corresponding polypeptide level. In another embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes and/or corresponding polypeptides.

As described and defined herein, soft tissue sarcoma cancer proteins, including intracellular, transmembrane, or secreted proteins, find use as markers of soft tissue sarcoma cancer, e.g., for prognostic or diagnostic purposes. Detection of these proteins in putative soft tissue sarcoma cancer tissue allows for detection, prognosis, or diagnosis of cancer, and for selection of therapeutic strategy. In one embodiment, antibodies are used to detect soft tissue sarcoma cancer proteins. One method

separates proteins from a sample by electrophoresis on a gel (typically a denaturing and reducing protein gel, but may be another type of gel, including isoelectric focusing gels and the like). Following separation of proteins, the cancer protein is detected, e.g., by immunoblotting with antibodies raised against the cancer protein.

In another method, antibodies to the soft tissue sarcoma cancer protein find use in in situ imaging techniques, e.g., in histology. See, e.g., Asai, et al. (eds. 1993) Methods in Cell Biology: Antibodies in Cell Biology (vol. 37) Academic Press. In this method cells are contacted with from one to many antibodies to the soft tissue sarcoma cancer protein(s). Following washing to remove non-specific antibody binding, the presence of the antibody or antibodies is detected. In one embodiment the antibody is detected by incubating with a secondary antibody that contains a detectable label. In another method the primary antibody to the soft tissue sarcoma cancer protein(s) contains a detectable label, e.g., an enzyme marker that can act on a substrate. In another embodiment each one of multiple primary antibodies contains a distinct and detectable label. This method finds particular use in simultaneous screening for a plurality of soft tissue sarcoma cancer proteins. Many other histological imaging techniques are also provided by the invention.

In one embodiment the label is detected in a fluorometer which has the ability to detect and distinguish emissions of different wavelengths. In addition, a fluorescence activated cell sorter (FACS) can be used in the method.

In another embodiment, antibodies find use in diagnosing soft tissue sarcoma cancer from biological samples, such as blood, urine, sputum, semen, or other bodily fluids. As previously described, certain cancer proteins are secreted/circulating molecules. Blood or semen samples, therefore, are useful as samples to be probed or tested for the presence of secreted cancer proteins. Antibodies can be used to detect a soft tissue sarcoma cancer protein by previously described immunoassay techniques including ELISA, immunoblotting (Western blotting), immunoprecipitation, BIACORE technology and the like. Conversely, the presence of antibodies may indicate an immune response against an endogenous soft tissue sarcoma cancer protein.

In one embodiment, in situ hybridization of labeled cancer nucleic acid probes to tissue arrays is done. For example, arrays of tissue samples, including cancer tissue and/or normal tissue, are made. In situ hybridization (see, e.g., Ausubel, supra) is then performed. When comparing the fingerprints between an individual and a

standard, the skilled artisan can make a diagnosis, a prognosis, or a prediction based on the findings. It is further understood that the genes which indicate the diagnosis may differ from those which indicate the prognosis and molecular profiling of the condition of the cells may lead to distinctions between responsive or refractory conditions or may be predictive of outcomes.

Assays for Prognosis of Soft Tissue Sarcoma Disorders

In one embodiment, the cancer proteins, antibodies, nucleic acids, modified proteins and cells containing cancer sequences are used in prognosis assays. As above, gene expression profiles can be generated that correlate to soft tissue sarcoma cancer severity, in terms of long term prognosis. Again, this may be done on either a protein or gene level, with the use of genes included. For example, P-glycoprotein and Ki-67 antigen are promising markers for 5-year overall and disease-free survival for soft tissue sarcoma patients. Levine E.A. et al., Evaluation of new prognostic markers for adult soft tissue sarcomas, J. Clin. Oncol. 15:3249-57 (1997). Other markers may be similarly

identified that correlate to soft tissue sarcoma cancer severity, or survival rates of soft tissue sarcoma patients.

As above, cancer probes may be attached to biochips for the detection and quantification of cancer sequences in a tissue or patient. The assays proceed as outlined above for diagnosis. PCR method may provide more sensitive and accurate quantification.

Genes useful in prognostic assays are genes that are differentially expressed according to the stage of illness of the patient. In one embodiment, the genes may be uniquely expressed according to the stage of the patient. In another embodiment, the genes may be expressed at differential levels according to the stage of the patient. INSERT EXAMPLE OF PROGNOSTIC ASSAY The correlation of genes expressed in the different stages, either uniquely expressed or have differential expression levels according to the stage, may be used to determine the viability of inducing remission in a patient. In addition, genes that are expressed indicating onset of long-term complications may also be useful as a prognostic tool.

Assays for therapeutic compounds

The soft tissue sarcoma cancer proteins, antibodies, nucleic acids, modified proteins and cells containing soft tissue sarcoma cancer sequences are used in drug screening assays or by evaluating the effect of drug candidates on a "gene expression profile" or expression profile of polypeptides. In one embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes after treatment with a candidate agent. See, e.g., Zlokarnik, et al. (1998) Science 279:84-88; and Heid (1996) Genome Res. 6:986-994.

In one embodiment, the soft tissue sarcoma cancer proteins, antibodies, nucleic acids, modified proteins and cells containing the native or modified soft tissue sarcoma cancer proteins are used in screening assays. That is, the present invention provides novel methods for screening for compositions which modulate the soft tissue sarcoma cancer phenotype or an identified physiological function of a soft tissue sarcoma cancer protein. As above, this can be done on an individual gene level or by evaluating the effect of drug candidates on a "gene expression profile". In one embodiment, the expression profiles are used, preferably in conjunction with high throughput screening techniques to allow monitoring for expression profile genes after treatment with a candidate agent, see Zlokarnik, supra.

Having identified the differentially expressed genes herein, a variety of assays may be executed. In one embodiment, assays may be run on an individual gene or protein level. That is, having identified a particular gene as up regulated in soft tissue sarcoma cancer, test compounds can be screened for the ability to modulate gene expression or for binding to the cancer protein. "Modulation" thus includes both an increase and a decrease in gene expression. The amount of modulation will depend on the original change of the gene expression in normal versus tissue exhibiting soft tissue sarcoma cancer, with changes of at least about 10%, preferably about 50%, more preferably about 100-300%, and in some embodiments 300-1000% or greater. Thus, if a gene exhibits a 4-fold increase in cancer tissue compared to normal tissue, a decrease of about four-fold is often desired; similarly, a 10-fold decrease in cancer tissue compared to normal tissue often provides a target value of a 10-fold increase in expression to be induced by the test compound.

The amount of gene expression may be monitored using nucleic acid probes and the quantification of gene expression levels, or, alternatively, the gene product

itself can be monitored, e.g., through the use of antibodies to the cancer protein and standard immunoassays. Proteomics and separation techniques may also allow quantification of expression.

In one embodiment, gene expression or protein monitoring of a number of entities, e.g., an expression profile, is monitored simultaneously. Such profiles will typically involve a plurality of those entities described herein.

In this embodiment, the soft tissue sarcoma cancer nucleic acid probes are attached to biochips as outlined herein for the detection and quantification of cancer sequences in a particular cell. Alternatively, PCR may be used. Thus, a series, e.g., of microtiter plate, may be used with dispensed primers in desired wells. A PCR reaction can then be performed and analyzed for each well.

Modulators of soft tissue sarcoma cancer

Expression monitoring can be performed to identify compounds that modify the expression of one or more soft tissue sarcoma cancer-associated sequences, e.g., a polynucleotide sequence set out in the Tables. Generally, in one embodiment, a test modulator is added to the cells prior to analysis. Moreover, screens are also provided to identify agents that modulate soft tissue sarcoma cancer, modulate soft tissue sarcoma cancer proteins, bind to a soft tissue sarcoma cancer protein, or interfere with the binding of a soft tissue sarcoma cancer protein and an antibody or other binding partner.

The term "test compound" or "drug candidate" or "modulator" or grammatical equivalents as used herein describes a molecule, e.g., protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide, etc., to be tested for the capacity to directly or indirectly alter the cancer phenotype or the expression of a cancer sequence, e.g., a nucleic acid or protein sequence. In several embodiments, modulators alter expression profiles, or expression profile nucleic acids or proteins provided herein. In one embodiment, the modulator suppresses a soft tissue sarcoma cancer phenotype, e.g., to a normal tissue fingerprint. In another embodiment, a modulator induced a cancer phenotype. Generally, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, e.g., at zero concentration or below the level of detection.

Drug candidates encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 daltons. Small molecules may be less than 2000, or less than 1500, or less than 1000, or less than 500 D. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs, or combinations thereof. Particularly useful are peptides.

In one aspect, a modulator will neutralize the effect of a soft tissue sarcoma cancer protein. By "neutralize" is meant that activity of a protein is inhibited or blocked and thereby has substantially no effect on a cell.

In certain embodiments, combinatorial libraries of potential modulators will be screened for an ability to bind to a soft tissue sarcoma cancer polypeptide or to modulate activity. Conventionally, new chemical entities with useful properties are generated by identifying a chemical compound (called a "lead compound") with some desirable property or activity, e.g., inhibiting activity, creating variants of the lead compound, and evaluating the property and activity of those variant compounds. Often, high throughput screening (HTS) methods are employed for such an analysis. See, e.g., Janzen (2002) High Throughput Screening: Methods and Protocols Humana; Devlin (ed. 1997) High Throughput Screening: The Discovery of Bioactive Substances Dekker; and Mei and Czarnik (eds. 2002) Integrated Drug Discovery Techniques Dekker.

In one embodiment, high throughput screening methods involve providing a library containing a large number of potential therapeutic compounds (candidate compounds). Such "combinatorial chemical libraries" are then screened in one or more assays to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library, such as a polypeptide (e.g., mutein) library, is formed by combining a set of chemical building blocks called amino acids in every possible way for a given compound length (e.g., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks. See Gallop, et al. (1994) J. Med. Chem. 37:1233-1251.

Preparation and screening of combinatorial chemical libraries is well known. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., US Patent No. 5,010,175, Furka (1991) Pept. Prot. Res. 37:487-493, Houghton, et al. (1991) Nature 354:84-88), peptoids (PCT Publication No WO 91/19735), encoded peptides (PCT Publication WO 93/20242), random bio-oligomers (PCT Publication WO 92/00091), benzodiazepines (US Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs, et al. (1993) Proc. Nat'l Acad. Sci. USA 90:6909-6913), vinylogous polypeptides (Hagihara, et al. (1992) J. Amer. Chem. Soc. 114:6568), nonpeptidal peptidomimetics with a Beta-D-Glucose scaffolding (Hirschmann, et al. (1992) J. Amer. Chem. Soc. 114:9217-9218), analogous organic syntheses of small compound libraries (Chen, et al. (1994) J. Amer. Chem. Soc. 116:2661), oligocarbamates (Cho, et al. (1993) Science 261:1303-1305), and/or peptidyl phosphonates (Campbell, et al. (1994) J. Org. Chem. 59:658-xxx). See, generally, Gordon, et al. (1994) J. Med. Chem. 37:1385-1401, nucleic acid libraries (see, e.g., Stratagene, Corp.), peptide nucleic acid libraries (see, e.g., US Patent 5,539,083), antibody libraries (see, e.g., Vaughn, et al. (1996) Nature Biotechnology 14:309-314, and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang, et al. (1996) Science 274:1520-1522, and US Patent No. 5,593,853), and small organic molecule libraries (see, e.g., benzodiazepines, page 33 Baum (Jan 18, 1993) C&E News); isoprenoids (US Patent No. 5,569,588); thiazolidinones and metathiazanones (US Patent No. 5,549,974); pyrrolidines (US Patent Nos. 5,525,735 and 5,519,134); morpholino compounds (US Patent No. 5,506,337); benzodiazepines (US Patent No. 5,288,514); and the like.

Devices for the preparation of combinatorial libraries are commercially available. See, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY;

Symphony, Rainin, Woburn, MA; 433A Applied Biosystems, Foster City, CA; 9050 Plus, Millipore, Bedford, MA.

A number of well known robotic systems have also been developed for solution phase chemistries. These systems include automated workstations like the automated synthesis apparatus developed by Takeda Chemical Industries, LTD. (Osaka, Japan) and many robotic systems utilizing robotic arms (Zymate II, Zymark Corporation, Hopkinton, MA; Orca, Hewlett-Packard, Palo Alto, CA), which mimic the manual synthetic operations performed by a chemist. The above devices are suitable for use with the present invention. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent. In addition, numerous combinatorial libraries are themselves commercially available. See, e.g., ComGenex, Princeton, NJ; Asinex, Moscow, Ru; Tripos, Inc., St. Louis, MO; ChemStar, Ltd, Moscow, RU; 3D Pharmaceuticals, Exton, PA; Martek Biosciences, Columbia, MD; etc.

The assays to identify modulators are amenable to high throughput screening. Assays thus detect enhancement or inhibition of cancer gene transcription, inhibition or enhancement of polypeptide expression, and inhibition or enhancement of polypeptide activity.

High throughput assays for the presence, absence, quantification, or other properties of particular nucleic acids or protein products are well known, as are binding assays and reporter gene assays. Thus, e.g., US Patent No. 5,559,410 discloses high throughput screening methods for proteins, US Patent No. 5,585,639 discloses high throughput screening methods for nucleic acid binding (e.g., in arrays), while US Patent Nos. 5,576,220 and 5,541,061 disclose high throughput methods of screening for ligand/antibody binding.

In addition, high throughput screening systems are commercially available. See, e.g., Zymark Corp., Hopkinton, MA; Air Technical Industries, Mentor, OH; Beckman Instruments, Inc., Fullerton, CA; Precision Systems, Inc., Natick, MA, etc. These systems typically automate entire procedures, including sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. The manufacturers of such systems provide detailed protocols for various high throughput systems. Thus, e.g., Zymark Corp. provides technical bulletins describing

screening systems for detecting the modulation of gene transcription, ligand binding, and the like.

In one embodiment, modulators are proteins, often naturally occurring proteins or fragments of naturally occurring proteins. Thus, e.g., cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of proteins may be made for screening in the methods of the invention. Particularly useful in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, including human proteins. Particularly useful test compound will be directed to the class of proteins to which the target belongs, e.g., substrates for enzymes or ligands and receptors.

In one embodiment, modulators are peptides of from about 5-30 amino acids, with from about 5-20 amino acids being preferred, and from about 7-15 being particularly preferred. The peptides may be digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate nucleotide or amino acid variations. The synthetic process can be designed to generate randomized proteins or nucleic acids, to allow the formation of the possible combinations over the length of the sequence, thus forming a library of randomized candidate bioactive proteinaceous agents.

In one embodiment, the library is fully randomized, with no sequence preferences or constants. In one embodiment, the library is biased. That is, some positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in one embodiment, the nucleotides or amino acid residues are randomized within a defined class, e.g., of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of nucleic acid binding domains, the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines, or histidines for phosphorylation sites, etc., or to purines, etc.

Modulators of soft tissue sarcoma cancer can also be nucleic acids, as defined above.

As described above generally for proteins, nucleic acid modulating agents may be naturally occurring nucleic acids, random nucleic acids, or "biased" random

nucleic acids. For example, digests of prokaryotic or eukaryotic genomes may be used as is outlined above for proteins.

In one embodiment, the candidate compounds are organic chemical moieties, a wide variety of which are available in the literature.

After the candidate agent has been added and the cells allowed to incubate for some period of time, the sample containing a target sequence to be analyzed is added to the biochip. If required, the target sequence is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, electroporation, etc., with purification and/or amplification such as PCR performed as appropriate. For example, an in vitro transcription with labels covalently attached to the nucleotides is performed. Generally, the nucleic acids are labeled with biotin-FITC or PE, or with cy3 or cy5.

In one embodiment, the target sequence is labeled with, e.g., a fluorescent, a chemiluminescent, a chemical, or a radioactive signal, to provide a means of detecting the target sequence's specific binding to a probe. The label also can be an enzyme, such as, alkaline phosphatase or horseradish peroxidase, which when provided with an appropriate substrate produces a product that can be detected. Alternatively, the label can be a labeled compound or small molecule, such as an enzyme inhibitor, that binds but is not catalyzed or altered by the enzyme. The label also can be a moiety or compound, such as, an epitope tag or biotin which specifically binds to streptavidin. For the example of biotin, the streptavidin is labeled as described above, thereby, providing a detectable signal for the bound target sequence. Unbound labeled streptavidin is typically removed prior to analysis.

These assays can be direct hybridization assays or can comprise "sandwich assays", which include the use of multiple probes. See, e.g., US Patent Nos. 5,681,702; 5,597,909; 5,545,730; 5,594,117; 5,591,584; 5,571,670; 5,580,731; 5,571,670; 5,591,584; 5,624,802; 5,635,352; 5,594,118; 5,359,100; 5,124,246; and 5,681,697. The target nucleic acid may be prepared as outlined above, and then added to the biochip comprising a plurality of nucleic acid probes, under conditions that allow the formation of a hybridization complex.

A variety of hybridization conditions may be used in the present invention, including high, moderate, and low stringency conditions as outlined above. The assays are generally run under stringency conditions which allows formation of the label probe hybridization complex only in the presence of target. Stringency can be

controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration, pH, organic solvent concentration, etc.

These parameters may also be used to control non-specific binding. See US Patent No. 5,681,697. Thus it may be desirable to perform certain steps at higher stringency conditions to reduce non-specific binding.

The reactions outlined herein may be accomplished in a variety of ways. Components of the reaction may be added simultaneously, or sequentially, in different orders, with several embodiments outlined below. In addition, the reaction may include a variety of other reagents. These include salts, buffers, neutral proteins, e.g., albumin, detergents, etc., which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may also be used as appropriate, depending on the sample preparation methods and purity of the target.

The assay data are analyzed to determine the expression levels, and changes in expression levels as between states, of individual genes, forming a gene expression profile.

Screens are performed to identify modulators of the soft tissue sarcoma cancer phenotype. In one embodiment, screening is performed to identify modulators that can induce or suppress a particular expression profile, thus preferably generating the associated phenotype. In another embodiment, e.g., for diagnostic applications, having identified differentially expressed genes important in a particular state, screens can be performed to identify modulators that alter expression of individual genes. In an another embodiment, screening is performed to identify modulators that alter a biological function of the expression product of a differentially expressed gene. Again, having identified the importance of a gene in a particular state, screens are performed to identify agents that bind and/or modulate the biological activity of the gene product.

In addition, screens can be done for genes that are induced in response to a candidate agent. After identifying a modulator based upon its ability to suppress a soft tissue sarcoma cancer expression pattern leading to a normal expression pattern, or to modulate a single cancer gene expression profile so as to mimic the expression of the gene from normal tissue, a screen as described above can be performed to

identify genes that are specifically modulated in response to the agent. Comparing expression profiles between normal tissue and agent treated soft tissue sarcoma cancer tissue reveals genes that are not expressed in normal tissue or soft tissue sarcoma cancer tissue, but are expressed in agent treated tissue. These agent-specific sequences can be identified and used by methods described herein for cancer genes or proteins. In particular, these sequences and the proteins they encode find use in marking or identifying agent treated cells. In addition, antibodies can be raised against the agent induced proteins and used to target novel therapeutics to the treated soft tissue sarcoma cancer tissue sample.

Thus, in one embodiment, a test compound is administered to a population of cancer cells, that have an associated soft tissue sarcoma cancer expression profile. By "administration" or "contacting" herein is meant that the candidate agent is added to the cells in such a manner as to allow the agent to act upon the cell, whether by uptake and intracellular action, or by action at the cell surface. In some embodiments, nucleic acid encoding a proteinaceous candidate agent (e.g., a peptide) may be put into a viral construct such as an adenoviral or retroviral construct, and added to the cell, such that expression of the peptide agent is achieved. See PCT US97/01019. Regulatable gene therapy systems can also be used.

Once the test compound has been administered to the cells, the cells can be washed if desired and are allowed to incubate under preferably physiological conditions for some period of time. The cells are then harvested and a new gene expression profile is generated, as outlined herein.

Thus, e.g., soft tissue sarcoma cancer tissue may be screened for agents that modulate, e.g., induce or suppress the cancer phenotype. A change in at least one gene, preferably many, of the expression profile indicates that the agent has an effect on the cancer activity. By defining such a signature for the soft tissue sarcoma cancer phenotype, screens for new drugs that alter the phenotype can be devised. With this approach, the drug target need not be known and need not be represented in the original expression screening platform, nor does the level of transcript for the target protein need to change.

In one embodiment, as outlined above, screens may be done on individual genes and gene products (proteins). That is, having identified a particular differentially expressed gene as important in a particular state, screening of modulators of either the expression of the gene or the gene product itself can be done.

The gene products of differentially expressed genes are sometimes referred to herein as "cancer proteins" or a "cancer modulatory protein". The cancer modulatory protein may be a fragment, or alternatively, be the full length protein to the fragment encoded by the nucleic acids of the Tables. Preferably, the cancer modulatory protein is a fragment. In one embodiment, the cancer amino acid sequence which is used to determine sequence identity or similarity is encoded by a nucleic acid of the Tables. In another embodiment, the sequences are naturally occurring allelic variants of a protein encoded by a nucleic acid of the Tables. In another embodiment, the sequences are sequence variants as further described herein.

Preferably, the cancer modulatory protein is a fragment of approximately 14-24 amino acids long. More preferably the fragment is a soluble fragment. Preferably, the fragment includes a non-transmembrane region. In one embodiment, the fragment has an N-terminal Cys to aid in solubility. In one embodiment, the C-terminus of the fragment is kept as a free acid and the N-terminus is a free amine to aid in coupling, e.g., to cysteine.

In one embodiment the cancer proteins are conjugated to an immunogenic agent as discussed herein. In one embodiment the cancer protein is conjugated to BSA

Measurements of soft tissue sarcoma cancer polypeptide activity, or of soft tissue sarcoma cancer or cancer phenotype can be performed using a variety of assays. For example, the effects of the test compounds upon the function of the soft tissue sarcoma cancer polypeptides can be measured by examining parameters described above. A suitable physiological change that affects activity can be used to assess the influence of a test compound on the polypeptides of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects associated with tumors, tumor growth, neovascularization, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP. In the assays of the invention, mammalian soft tissue sarcoma cancer polypeptide is typically used, e.g., mouse, preferably human.

Assays to identify compounds with modulating activity can be performed in vitro. For example, a soft tissue sarcoma cancer polypeptide is first contacted with a potential modulator and incubated for a suitable amount of time, e.g., from about 0.5-

48 hours. In one embodiment, the soft tissue sarcoma cancer polypeptide levels are determined in vitro by measuring the level of protein or mRNA. The level of protein is typically measured using immunoassays such as western blotting, ELISA, and the like with an antibody that selectively binds to the soft tissue sarcoma cancer polypeptide or a fragment thereof. For measurement of mRNA, amplification, e.g., using PCR, LCR, or hybridization assays, e.g., northern hybridization, RNAse protection, dot blotting, are included. The level of protein or mRNA is typically detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using a soft tissue sarcoma cancer protein promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or β-gal. The reporter construct is typically transfected into a cell. After treatment with a potential modulator, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques.

In one embodiment, as outlined above, screens may be done on individual genes and gene products (proteins). That is, having identified a particular differentially expressed gene as important in a particular state, screening of modulators of the expression of the gene or the gene product itself can be done. The gene products of differentially expressed genes are sometimes referred to herein as "soft tissue sarcoma cancer proteins." The soft tissue sarcoma cancer protein may be a fragment, or alternatively, the full length protein to a fragment shown herein.

In one embodiment, screening for modulators of expression of specific genes is performed. Typically, the expression of only one or a few genes are evaluated. In another embodiment, screens are designed to first find compounds that bind to differentially expressed proteins. These compounds are then evaluated for the ability to modulate differentially expressed activity. Moreover, once initial candidate compounds are identified, variants can be further screened to better evaluate structure activity relationships.

In another embodiment, binding assays are done. In general, purified or isolated gene product is used; that is, the gene products of one or more differentially expressed nucleic acids are made. For example, antibodies are generated to the protein gene products, and standard immunoassays are run to determine the amount of

protein present. Alternatively, cells comprising the soft tissue sarcoma cancer proteins can be used in the assays.

Thus, in one embodiment, the methods comprise combining a soft tissue sarcoma cancer protein and a candidate compound, and determining the binding of the compound to the soft tissue sarcoma cancer protein. Other embodiments utilize the human soft tissue sarcoma cancer protein, although other mammalian proteins may also be used, e.g., for the development of animal models of human disease. In some embodiments, as outlined herein, variant or derivative soft tissue sarcoma cancer proteins may be used.

Generally, in one embodiment of the methods herein, the soft tissue sarcoma cancer protein or the candidate agent is non-diffusibly bound to an insoluble support, preferably having isolated sample receiving areas (e.g., a microtiter plate, an array, etc.). The insoluble supports may be made of a composition to which the compositions can be bound, is readily separated from soluble material, and is otherwise compatible with the overall method of screening. The surface of such supports may be solid or porous and of a convenient shape. Examples of suitable insoluble supports include microtiter plates, arrays, membranes and beads. These are typically made of glass, plastic (e.g., polystyrene), polysaccharides, nylon, or nitrocellulose, TEFLON® (synthetic resinous flurorine-containing polymers), etc. Microtiter plates and arrays are especially convenient because a large number of assays can be carried out simultaneously, using small amounts of reagents and samples. The particular manner of binding of the composition is typically compatible with the reagents and overall methods of the invention, maintains the activity of the composition, and is nondiffusable. Other methods of binding include the use of antibodies (which do not sterically block either the ligand binding site or activation sequence when the protein is bound to the support), direct binding to "sticky" or ionic supports, chemical crosslinking, the synthesis of the protein or agent on the surface, etc. Following binding of the protein or agent, excess unbound material is removed by washing. The sample receiving areas may then be blocked through incubation with bovine serum albumin (BSA), casein, or other innocuous protein or other moiety.

In one embodiment, the soft tissue sarcoma cancer protein is bound to the support, and a test compound is added to the assay. Alternatively, the candidate agent is bound to the support and the cancer protein is added. Novel binding agents include

specific antibodies, non-natural binding agents identified in screens of chemical libraries, peptide analogs, etc. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, functional assays (phosphorylation assays, etc.), and the like.

The determination of the binding of the test modulating compound to the cancer protein may be done in many ways. In one embodiment, the compound is labeled, and binding determined directly, e.g., by attaching all or a portion of the cancer protein to a solid support, adding a labeled candidate agent (e.g., a fluorescent label), washing off excess reagent, and determining whether the label is present on the solid support. Various blocking and washing steps may be utilized as appropriate.

In some embodiments, only one of the components is labeled, e.g., the proteins (or proteinaceous candidate compounds) can be labeled. Alternatively, more than one component can be labeled with different labels, e.g., ^{125}I for the proteins and a fluorophor for the compound. Proximity reagents, e.g., quenching or energy transfer reagents are also useful.

In one embodiment, the binding of the test compound is determined by competitive binding assay. The competitor may be a binding moiety known to bind to the target molecule (e.g., a soft tissue sarcoma cancer protein), such as an antibody, peptide, binding partner, ligand, etc. Under certain circumstances, there may be competitive binding between the compound and the binding moiety, with the binding moiety displacing the compound. In one embodiment, the test compound is labeled. Either the compound, or the competitor, or both, is added first to the protein for a time sufficient to allow binding, if present. Incubations may be performed at a temperature which facilitates optimal activity, typically between about 4-40° C. Incubation periods are typically optimized, e.g., to facilitate rapid high throughput screening. Typically between about 0.1-1 hour will be sufficient. Excess reagent is generally removed or washed away. The second component is then added, and the presence or absence of the labeled component is followed, to indicate binding.

In one embodiment, the competitor is added first, followed by a test compound. Displacement of the competitor is an indication that the test compound is binding to the cancer protein and thus is capable of binding to, and potentially

modulating, the activity of the cancer protein. In this embodiment, either component can be labeled. Thus, e.g., if the competitor is labeled, the presence of label in the wash solution indicates displacement by the agent. Alternatively, if the test compound is labeled, the presence of the label on the support indicates displacement.

In an alternative embodiment, the test compound is added first, with incubation and washing, followed by the competitor. The absence of binding by the competitor may indicate that the test compound is bound to the cancer protein with a higher affinity. Thus, if the test compound is labeled, the presence of the label on the support, coupled with a lack of competitor binding, may indicate that the test compound is capable of binding to the cancer protein.

In one embodiment, the methods comprise differential screening to identity agents that are capable of modulating the activity of the cancer proteins. In one embodiment, the methods comprise combining a cancer protein and a competitor in a first sample. A second sample comprises a test compound, a cancer protein, and a competitor. The binding of the competitor is determined for both samples, and a change, or difference in binding between the two samples indicates the presence of an agent capable of binding to the cancer protein and potentially modulating its activity. That is, if the binding of the competitor is different in the second sample relative to the first sample, the agent is capable of binding to the cancer protein.

Alternatively, differential screening is used to identify drug candidates that bind to the native soft tissue sarcoma cancer protein, but cannot bind to modified soft tissue sarcoma cancer proteins. The structure of the cancer protein may be modeled, and used in rational drug design to synthesize agents that interact with that site. Drug candidates that affect the activity of the cancer protein are also identified by screening drugs for the ability to either enhance or reduce the activity of the protein.

Positive controls and negative controls may be used in the assays. Preferably control and test samples are performed in at least triplicate to obtain statistically significant results. Incubation of all samples is for a time sufficient for the binding of the agent to the protein. Following incubation, samples are washed free of non-specifically bound material and the amount of bound, generally labeled agent determined. For example, where a radiolabel is employed, the samples may be counted in a scintillation counter to determine the amount of bound compound.

A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g., albumin, detergents, etc., which may

be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in an order that provides for the requisite binding.

In one embodiment, the invention provides methods for screening for a compound capable of modulating the activity of a cancer protein. The methods comprise adding a test compound, as defined above, to a cell comprising cancer proteins. Useful cell types include almost any cell. The cells contain a recombinant nucleic acid that encodes a cancer protein. In one embodiment, a library of candidate agents are tested on a plurality of cells.

In one aspect, the assays are evaluated in the presence or absence or previous or subsequent exposure of physiological signals, e.g., hormones, antibodies, peptides, antigens, cytokines, growth factors, action potentials, pharmacological agents including chemotherapeutics, radiation, carcinogenics, or other cells (e.g., cell-cell contacts). In another example, the determinations are determined at different stages of the cell cycle process.

In this way, compounds that modulate soft tissue sarcoma cancer agents are identified. Compounds with pharmacological activity are able to enhance or interfere with the activity of the cancer protein. Once identified, similar structures are evaluated to identify critical structural feature of the compound.

In one embodiment, a method of inhibiting soft tissue sarcoma cancer cell division is provided. The method comprises administration of a soft tissue sarcoma cancer inhibitor. In another embodiment, a method of inhibiting soft tissue sarcoma cancer is provided. The method may comprise administration of a soft tissue sarcoma cancer inhibitor. In a further embodiment, methods of treating cells or individuals with soft tissue sarcoma cancer are provided, e.g., comprising administration of a soft tissue sarcoma cancer inhibitor.

In one embodiment, a soft tissue sarcoma cancer inhibitor is an antibody as discussed above. In another embodiment, the soft tissue sarcoma cancer inhibitor is an antisense molecule.

A variety of cell growth, proliferation, viability, and metastasis assays are known, as described below.

Soft agar growth or colony formation in suspension

Normal cells require a solid substrate to attach and grow. When the cells are transformed, they lose this phenotype and grow detached from the substrate. For example, transformed cells can grow in stirred suspension culture or suspended in semi-solid media, such as semi-solid or soft agar. The transformed cells, when transfected with tumor suppressor genes, regenerate normal phenotype and require a solid substrate to attach and grow. Soft agar growth or colony formation in suspension assays can be used to identify modulators of soft tissue sarcoma cancer sequences, which when expressed in host cells, inhibit abnormal cellular proliferation and transformation. A therapeutic compound would reduce or eliminate the host cells' ability to grow in stirred suspension culture or suspended in semi-solid media, such as semi-solid or soft.

Techniques for soft agar growth or colony formation in suspension assays are described, e.g., in Freshney (1998) Culture of Animal Cells: A Manual of Basic Technique (3d ed.) Wiley-Liss; Freshney (2000) Culture of Animal Cells: A Manual of Basic Technique (4th ed.) Wiley-Liss; and Garkavtsev, et al. (1996) Nature Genet. 14:415-20.

Contact inhibition and density limitation of growth

Normal cells typically grow in a flat and organized pattern in a petri dish until they touch other cells. When the cells touch one another, they are contact inhibited and stop growing. When cells are transformed, however, the cells are not contact inhibited and continue to grow to high densities in disorganized foci. Thus, the transformed cells grow to a higher saturation density than normal cells. This can be detected morphologically by the formation of a disoriented monolayer of cells or rounded cells in foci within the regular pattern of normal surrounding cells. Alternatively, labeling index with (^{3}H)-thymidine at saturation density can be used to measure density limitation of growth. See Freshney (2000), supra. The transformed cells, when transfected with tumor suppressor genes, regenerate a normal phenotype and become contact inhibited and would grow to a lower density.

In this assay, labeling index with (^{3}H)-thymidine at saturation density is a method of measuring density limitation of growth. Transformed host cells are transfected with a soft tissue sarcoma cancer-associated sequence and are grown for

24 hours at saturation density in non-limiting medium conditions. The percentage of cells labeling with (^{3}H)-thymidine is determined autoradiographically. See, Freshney (1998), supra.

Growth factor or serum dependence

Transformed cells typically have a lower serum dependence than their normal counterparts. See, e.g., Temin (1966) J. Nat'l Cancer Inst. 37:167-175; Eagle, et al. (1970) J. Exp. Med. 131:836-879; Freshney, supra. This is in part due to release of various growth factors by the transformed cells. Growth factor or serum dependence of transformed host cells can be compared with that of control.

Tumor specific markers levels

Tumor cells release an increased amount of certain factors (hereinafter "tumor specific markers") than their normal counterparts. For example, plasminogen activator (PA) is released from human glioma at a higher level than from normal brain cells. See, e.g., Gullino "Angiogenesis, tumor vascularization, and potential interference with tumor growth" pp. 178-184 in Mihich (ed. 1985) Biological Responses in Cancer Plenum. Similarly, tumor angiogenesis factor (TAF) is released at a higher level in tumor cells than their normal counterparts. See, e.g., Folkman (1992) "Angiogenesis and Cancer" Sem Cancer Biol. 3:89-96.

Various techniques which measure the release of these factors are described in Freshney (1994), supra. See also, Unkeless, et al. (1974) J. Biol. Chem. 249:4295-4305; Strickland and Beers (1976) J. Biol. Chem. 251:5694-5702; Whur, et al. (1980) Br. J. Cancer 42:305-312; Gullino, "Angiogenesis, tumor vascularization, and potential interference with tumor growth" pp. 178-184 in Mihich (ed. 1985) Biological Responses in Cancer Plenum; Freshney (1985) Anticancer Res. 5:111-130.

Invasiveness into Matrigel

The degree of invasiveness into Matrigel or some other extracellular matrix constituent can be used as an assay to identify compounds that modulate soft tissue sarcoma cancer-associated sequences. Tumor cells exhibit a good correlation between malignancy and invasiveness of cells into Matrigel or some other extracellular matrix constituent. In this assay, tumorigenic cells are typically used as

host cells. Expression of a tumor suppressor gene in these host cells would decrease invasiveness of the host cells.

Techniques described in Freshney (1994), supra, can be used. Briefly, the level of invasion of host cells can be measured by using filters coated with Matrigel or some other extracellular matrix constituent. Penetration into the gel, or through to the distal side of the filter, is rated as invasiveness, and rated histologically by number of cells and distance moved, or by prelabeling the cells with ^{125}I and counting the radioactivity on the distal side of the filter or bottom of the dish. See, e.g., Freshney (1984), supra.

Tumor growth in vivo

Effects of soft tissue sarcoma cancer-associated sequences on cell growth can be tested in transgenic or immune-suppressed mice. Knock-out transgenic mice can be made, in which the cancer gene is disrupted or in which a cancer gene is inserted. Knock-out transgenic mice can be made by insertion of a marker gene or other heterologous gene into the endogenous cancer gene site in the mouse genome via homologous recombination. Such mice can also be made by substituting the endogenous cancer gene with a mutated version of the cancer gene, or by mutating the endogenous cancer gene, e.g., by exposure to carcinogens.

A DNA construct is introduced into the nuclei of embryonic stem cells. Cells containing the newly engineered genetic lesion are injected into a host mouse embryo, which is re-implanted into a recipient female. Some of these embryos develop into chimeric mice that possess germ cells partially derived from the mutant cell line. Therefore, by breeding the chimeric mice it is possible to obtain a new line of mice containing the introduced genetic lesion. See, e.g., Capecchi, et al. (1989) Science 244:1288-1292. Chimeric targeted mice can be derived according to Hogan, et al. (1988) Manipulating the Mouse Embryo: A Laboratory Manual CSH Press; and Robertson (ed. 1987) Teratocarcinomas and Embryonic Stem Cells: A Practical Approach IRL Press, Washington, D.C.

Alternatively, various immune-suppressed or immune-deficient host animals can be used. For example, genetically athymic "nude" mouse (see, e.g., Giovanella, et al. (1974) J. Natl. Cancer Inst. 52:921-930), a SCID mouse, a thymectomized mouse, or an irradiated mouse (see, e.g., Bradley, et al. (1978) Br. J. Cancer 38:263-272;

Selby, et al. (1980) Br. J. Cancer 41:52-61) can be used as a host. Transplantable tumor cells (typically about 10^6 cells) injected into isogenic hosts will produce invasive tumors in a high proportions of cases, while normal cells of similar origin will not. In hosts which developed invasive tumors, cells expressing a soft tissue sarcoma cancer-associated sequences are injected subcutaneously. After a suitable length of time, preferably about 4-8 weeks, tumor growth is measured (e.g., by volume or by its two largest dimensions) and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth.

Polynucleotide modulators of soft tissue sarcoma cancer

Antisense and RNAi Polynucleotides

In certain embodiments, the activity of a soft tissue sarcoma cancer-associated protein is downregulated, or entirely inhibited, by the use of antisense polynucleotide, e.g., a nucleic acid complementary to, and which can preferably hybridize specifically to, a coding mRNA nucleic acid sequence, e.g., a soft tissue sarcoma cancer protein mRNA, or a subsequence thereof. Binding of the antisense polynucleotide to the mRNA reduces the translation and/or stability of the mRNA.

In the context of this invention, antisense polynucleotides can comprise naturally-occurring nucleotides, or synthetic species formed from naturally-occurring subunits or their close homologs. Antisense polynucleotides may also have altered sugar moieties or inter-sugar linkages. Exemplary among these are the phosphorothioate and other sulfur containing species. Analogs are comprehended by this invention so long as they function effectively to hybridize with the soft tissue sarcoma cancer protein mRNA. See, e.g., Isis Pharmaceuticals, Carlsbad, CA; Sequitor, Inc., Natick, MA.

Such antisense polynucleotides can readily be synthesized using recombinant means, or can be synthesized in vitro. Equipment for such synthesis is sold by several vendors, including Applied Biosystems. The preparation of other oligonucleotides such as phosphorothioates and alkylated derivatives is also well known.

Antisense molecules as used herein include antisense or sense oligonucleotides. Sense oligonucleotides can, e.g., be employed to block transcription by binding to the anti-sense strand. The antisense and sense oligonucleotide comprise

a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target mRNA (sense) or DNA (antisense) sequences for soft tissue sarcoma cancer molecules. An antisense molecule is for a soft tissue sarcoma cancer sequence in Tables 1A-11C, or for a ligand or activator thereof. Antisense or sense oligonucleotides, according to the present invention, comprise a fragment generally at least about 14 nucleotides, preferably from about 14 to 30 nucleotides. The ability to derive an antisense or a sense oligonucleotide, based upon a cDNA sequence encoding a given protein is described in, e.g., Stein and Cohen (1988) Cancer Res. 48:2659-2668; and van der Krol, et al. (1988) BioTechniques 6:958-976.

RNA interference is a mechanism to suppress gene expression in a sequence specific manner. See, e.g., Brumelkamp, et al. (2002) Sciencexpress (21March2002); Sharp (1999) Genes Dev. 13:139-141; and Cathew (2001) Curr. Op. Cell Biol. 13:244-248. In mammalian cells, short, e.g., 21 nt, double stranded small interfering RNAs (siRNA) have been shown to be effective at inducing an RNAi response. See, e.g., Elbashir, et al. (2001) Nature 411:494-498. The mechanism may be used to downregulate expression levels of identified genes, e.g., treatment of or validation of relevance to disease.

Ribozymes

In addition to antisense polynucleotides, ribozymes can be used to target and inhibit transcription of soft tissue sarcoma cancer-associated nucleotide sequences. A ribozyme is an RNA molecule that catalytically cleaves other RNA molecules. Different kinds of ribozymes have been described, including group I ribozymes, hammerhead ribozymes, hairpin ribozymes, RNase P, and axhead ribozymes. See, e.g., Castanotto, et al. (1994) Adv. in Pharmacology 25:289-317.

General features of hairpin ribozymes are described, e.g., in Hampel, et al. (1990) Nuc. Acids Res. 18:299-304; European Patent Publication No. 0 360 257; US Patent No. 5,254,678. Methods of preparation are available. See, e.g., WO 94/26877; Yu, et al. (1993) Proc. Nat'l Acad. Sci. USA 90:6340-6344; Yamada, et al. (1994) Human Gene Therapy 1:39-45; Leavitt, et al. (1995) Proc. Nat'l Acad. Sci. USA 92:699-703; Leavitt, et al. (1994) Human Gene Therapy 5:1151-120; and Yamada, et al. (1994) Virology 205: 121-126.

Polynucleotide modulators of soft tissue sarcoma cancer may be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with

a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors. Preferably, conjugation of the ligand binding molecule does not substantially interfere with the ability of the ligand binding molecule to bind to its corresponding molecule or receptor, or block entry of the sense or antisense oligonucleotide or its conjugated version into the cell. Alternatively, a polynucleotide modulator of soft tissue sarcoma cancer may be introduced into a cell containing the target nucleic acid sequence, e.g., by formation of an polynucleotide-lipid complex, as described in WO 90/10448. It is understood that the use of antisense molecules or knock out and knock in models may also be used in screening assays as discussed above, in addition to methods of treatment.

Thus, in one embodiment, methods of modulating soft tissue sarcoma cancer in cells or organisms are provided. In one embodiment, the methods comprise administering to a cell an anti-soft tissue sarcoma cancer antibody that reduces or eliminates the biological activity of an endogenous soft tissue sarcoma cancer protein. Alternatively, the methods comprise administering to a cell or organism a recombinant nucleic acid encoding a soft tissue sarcoma cancer protein. This may be accomplished in many ways. In one embodiment, e.g., when the soft tissue sarcoma cancer sequence is down-regulated in soft tissue sarcoma cancer, such state may be reversed by increasing the amount of soft tissue sarcoma cancer gene product in the cell. This can be accomplished, e.g., by overexpressing the endogenous soft tissue sarcoma cancer gene or administering a gene encoding the soft tissue sarcoma cancer sequence, using known gene-therapy techniques. In one embodiment, the gene therapy techniques include the incorporation of the exogenous gene using enhanced homologous recombination (EHR), e.g., as described in PCT/US93/03868, hereby incorporated by reference in its entirety. Alternatively, e.g., when the soft tissue sarcoma cancer sequence is up-regulated in soft tissue sarcoma cancer, the activity of the endogenous soft tissue sarcoma cancer gene is decreased, e.g., by the administration of a soft tissue sarcoma cancer antisense nucleic acid or other inhibitor, such as RNAi.

In one embodiment, the soft tissue sarcoma cancer proteins of the present invention may be used to generate polyclonal and monoclonal antibodies to soft tissue sarcoma cancer proteins. Similarly, the soft tissue sarcoma cancer proteins can be coupled, using standard technology, to affinity chromatography columns. These

columns may then be used to purify cancer antibodies useful for production, diagnostic, or therapeutic purposes. In one embodiment, the antibodies are generated to epitopes unique to a cancer protein; that is, the antibodies show little or no cross-reactivity to other proteins. The soft tissue sarcoma cancer antibodies may be coupled to standard affinity chromatography columns and used to purify cancer proteins. The antibodies may also be used as blocking polypeptides, as outlined above, since they will specifically bind to the cancer protein.

Methods of identifying variant soft tissue sarcoma cancer-associated sequences

Without being bound by theory, expression of various soft tissue sarcoma cancer sequences is correlated with cancer. Accordingly, disorders based on mutant or variant soft tissue sarcoma cancer genes may be determined. In one embodiment, the invention provides methods for identifying cells containing variant soft tissue sarcoma cancer genes, e.g., determining all or part of the sequence of at least one endogenous soft tissue sarcoma cancer gene in a cell. This may be accomplished using known sequencing techniques. In one embodiment, the invention provides methods of identifying the cancer genotype of an individual, e.g., determining all or part of the sequence of at least one soft tissue sarcoma cancer gene of the individual. This is generally done in at least one tissue of the individual, and may include the evaluation of a number of tissues or different samples of the same tissue. The method may include comparing the sequence of the sequenced soft tissue sarcoma cancer gene to a known soft tissue sarcoma cancer gene, e.g., a wild-type gene.

The sequence of all or part of the soft tissue sarcoma cancer gene can then be compared to the sequence of a known soft tissue sarcoma cancer gene to determine if differences exist. This can be done using known homology programs, such as Bestfit, etc. In another embodiment, the presence of a difference in the sequence between the soft tissue sarcoma cancer gene of the patient and the known soft tissue sarcoma cancer gene correlates with a disease state or a propensity for a disease state, as outlined herein.

In one embodiment, the soft tissue sarcoma cancer genes are used as probes to determine the number of copies of the soft tissue sarcoma cancer gene in the genome.

In another embodiment, the soft tissue sarcoma cancer genes are used as probes to determine the chromosomal localization of the soft tissue sarcoma cancer genes. Information such as chromosomal localization finds use in providing a

diagnosis or prognosis in particular when chromosomal abnormalities such as translocations, and the like are identified in the soft tissue sarcoma cancer gene locus.

Administration of pharmaceutical and vaccine compositions

In one embodiment, a therapeutically effective dose of a soft tissue sarcoma cancer protein or modulator thereof, is administered to a patient. By "therapeutically effective dose" herein is meant a dose that produces effects for which it is administered. The exact dose will depend on the purpose of the treatment, and can be ascertained using known techniques. See, e.g., Ansel, et al. (1999) Pharmaceutical Dosage Forms and Drug Delivery Lippincott; Lieberman (1992) Pharmaceutical Dosage Forms (vols. 1-3) Dekker, ISBN 0824770846, 082476918X, 0824712692, 0824716981; Lloyd (1999) The Art, Science and Technology of Pharmaceutical Compounding Amer. Pharmaceut. Assn.; and Pickar (1998) Dosage Calculations Thomson. Adjustments for soft tissue sarcoma cancer degradation, systemic versus localized delivery, and rate of new protease synthesis, as well as the age, body weight, general health, sex, diet, time of administration, drug interaction, and the severity of the condition may be necessary. U.S. Patent Application N. 09/687,576, further discloses the use of compositions and methods of diagnosis and treatment in cancer.

A "patient" for the purposes of the present invention includes both humans and other animals, particularly mammals. Thus the methods are applicable to both human therapy and veterinary applications. In one embodiment the patient is a mammal, preferably a primate, including humans.

The administration of the soft tissue sarcoma cancer proteins and modulators thereof of the present invention can be done in a variety of ways as discussed above, including, but not limited to, orally, subcutaneously, intravenously, intranasally, transdermally topically, intraperitoneally, intramuscularly, intrapulmonary, vaginally, rectally, or intraocularly. In some instances, e.g., in the treatment of wounds and inflammation, the soft tissue sarcoma cancer proteins and modulators may be directly applied as a solution, spray, or ointment.

The pharmaceutical compositions of the present invention comprise a soft tissue sarcoma cancer protein in a form suitable for administration to a patient. In one embodiment, the pharmaceutical compositions are in a water soluble form, such as being present as pharmaceutically acceptable salts, which is meant to include both acid and base addition salts. "Pharmaceutically acceptable acid addition salt" refers to

those salts that retain the biological effectiveness of the free bases and that are not biologically or otherwise undesirable, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like, and organic acids such as acetic acid, propionic acid, glycolic acid, pyruvic acid, oxalic acid, maleic acid, malonic acid, succinic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid, and the like. "Pharmaceutically acceptable base addition salts" include those derived from inorganic bases such as sodium, potassium, lithium, ammonium, calcium, magnesium, iron, zinc, copper, manganese, aluminum salts, and the like. Particularly useful are the ammonium, potassium, sodium, calcium, and magnesium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, and ethanolamine.

The pharmaceutical compositions may also include one or more of the following: carrier proteins such as serum albumin; buffers; fillers such as microcrystalline cellulose, lactose, corn and other starches; binding agents; sweeteners and other flavoring agents; coloring agents; and polyethylene glycol.

The pharmaceutical compositions can be administered in a variety of unit dosage forms depending upon the method of administration. For example, unit dosage forms suitable for oral administration include, but are not limited to, powder, tablets, pills, capsules, and lozenges. It is recognized that soft tissue sarcoma cancer protein modulators (e.g., antibodies, antisense constructs, ribozymes, small organic molecules, etc.) when administered orally, should be protected from digestion. This is typically accomplished either by complexing the molecule(s) with a composition to render it resistant to acidic and enzymatic hydrolysis, or by packaging the molecule(s) in an appropriately resistant carrier, such as a liposome or a protection barrier. Means of protecting agents from digestion are available.

The compositions for administration will commonly comprise a soft tissue sarcoma cancer protein modulator dissolved in a pharmaceutically acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers can be used, e.g., buffered saline and the like. These solutions are sterile and generally free of

undesirable matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, e.g., sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate, and the like. The concentration of active agent in these formulations can vary widely, and will be selected primarily based on fluid volumes, viscosities, body weight, and the like in accordance with the particular mode of administration selected and the patient's needs. See, e.g., (1980) Remington's Pharmaceutical Science (18th ed.) Mack; and Hardman and Limbird (eds. 2001) Goodman and Gilman: The Pharmacological Basis of Therapeutics (10th ed.) McGraw-Hill.

Thus, a typical pharmaceutical composition for intravenous administration would be about 0.1 to 10 mg per patient per day. Dosages from about 0.1 to 100 mg per patient per day may be used, particularly when the drug is administered to a secluded site and not into the blood stream, such as into a body cavity or into a lumen of an organ. Substantially higher dosages are possible in topical administration. Actual methods for preparing parenterally administrable compositions will be known or apparent.

The compositions containing modulators of soft tissue sarcoma cancer proteins can be administered for therapeutic or prophylactic treatments. In therapeutic applications, compositions are administered to a patient suffering from a disease (e.g., a cancer) in an amount sufficient to cure or at least partially arrest the disease and its complications. An amount adequate to accomplish this is defined as a "therapeutically effective dose." Amounts effective for this use will depend upon the severity of the disease and the general state of the patient's health. Single or multiple administrations of the compositions may be administered depending on the dosage and frequency as required and tolerated by the patient. The composition should provide a sufficient quantity of the agents of this invention to effectively treat the patient. An amount of modulator that is capable of preventing or slowing the development of soft tissue sarcoma cancer in a mammal is referred to as a "prophylactically effective dose." The particular dose required for a prophylactic treatment will depend upon the medical condition and history of the mammal, the particular stage or form of soft tissue sarcoma cancer disorder being prevented, as well as other factors such as age, weight, gender, administration route, efficiency, etc.

Such prophylactic treatments may be used, e.g., in a mammal who has previously had cancer to prevent a recurrence of the cancer, or in a mammal who is suspected of having a significant likelihood of developing cancer, based, at least in part, upon gene expression profiles. Vaccine strategies may be used, in either a DNA vaccine form, or protein vaccine.

It will be appreciated that the present soft tissue sarcoma cancer protein-modulating compounds can be administered alone or in combination with additional soft tissue sarcoma cancer modulating compounds or with other therapeutic agent, e.g., other anti-cancer agents or treatments.

In numerous embodiments, one or more nucleic acids, e.g., polynucleotides comprising nucleic acid sequences set forth in Tables 1A-11C, such as RNAi, antisense polynucleotides, or ribozymes, will be introduced into cells, in vitro or in vivo. The present invention provides methods, reagents, vectors, and cells useful for expression of soft tissue sarcoma cancer-associated polypeptides and nucleic acids using in vitro (cell-free), ex vivo, or in vivo (cell or organism-based) recombinant expression systems.

The particular procedure used to introduce the nucleic acids into a host cell for expression of a protein or nucleic acid is application specific. Many procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, spheroplasts, electroporation, liposomes, microinjection, plasma vectors, viral vectors, and other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA, or other foreign genetic material into a host cell. See, e.g., Berger and Kimmel (1987) Guide to Molecular Cloning Techniques from Methods in Enzymology (vol. 152) Academic Press; Ausubel, et al. (eds. 1999 and supplements) Current Protocols Lippincott; and Sambrook, et al. (2001) Molecular Cloning: A Laboratory Manual (3d ed., Vol. 1-3) CSH Press.

In one embodiment, soft tissue sarcoma cancer proteins and modulators are administered as therapeutic agents, and can be formulated as outlined above. Similarly, cancer genes (including both the full-length sequence, partial sequences, or regulatory sequences of the cancer coding regions) can be administered in a gene therapy application. These cancer genes can include inhibitory applications, e.g., inhibitory RNA, gene therapy (e.g., for incorporation into the genome), or antisense compositions.

Soft tissue sarcoma cancer polypeptides and polynucleotides can also be administered as vaccine compositions to stimulate HTL, CTL, and antibody responses. Such vaccine compositions can include, e.g., lipidated peptides (see, e.g., Vitiello, et al. (1995) J. Clin. Invest. 95:341-349), peptide compositions encapsulated in poly(DL-lactide-co-glycolide) ("PLG") microspheres (see, e.g., Eldridge, et al. (1991) Molec. Immunol. 28:287-294; Alonso, et al. (1994) Vaccine 12:299-306; Jones, et al. (1995) Vaccine 13:675-681), peptide compositions contained in immune stimulating complexes (ISCOMS) (see, e.g., Takahashi, et al. (1990) Nature 344:873-875; Hu, et al. (1998) Clin. Exp. Immunol. 113:235-243), multiple antigen peptide systems (MAPs) (see, e.g., Tam (1988) Proc. Nat'l Acad. Sci. USA 85:5409-5413; Tam (1996) J. Immunol. Meth. 196:17-32), peptides formulated as multivalent peptides; peptides for use in ballistic delivery systems, typically crystallized peptides, viral delivery vectors (Perkus, et al., p. 379, in Kaufmann (ed. 1996) Concepts in Vaccine Development de Gruyter; Chakrabarti, et al. (1986) Nature 320:535-537; Hu, et al. (1986) Nature 320:537-540; Kieny, et al. (1986) Bio/Technology 4:790-795; Top, et al. (1971) J. Infect. Dis. 124:148-154; Chanda, et al. (1990) Virology 175:535-547), particles of viral or synthetic origin (see, e.g., Kofler, et al. (1996) J. Immunol. Meth. 192:25-35; Eldridge, et al. (1993) Sem. Hematol. 30:16-24; Falo, et al. (1995) Nature Med. 1:649-653), adjuvants (Warren, et al. (1986) Ann. Rev. Immunol. 4:369-388; Gupta, et al. (1993) Vaccine 11:293-306), liposomes (Reddy, et al. (1992) J. Immunol. 148:1585-1589; Rock (1996) Immunol. Today 17:131-137), or naked or particle absorbed cDNA (Ulmer, et al. (1993) Science 259:1745-1749; Robinson, et al. (1993) Vaccine 11:957-960; Shiver, et al., p 423, in Kaufmann (ed. 1996) Concepts in Vaccine Development de Gruyter; Cease and Berzofsky (1994) Ann. Rev. Immunol. 12:923-989; and Eldridge, et al. (1993) Sem. Hematol. 30:16-24). Toxin-targeted delivery technologies, also known as receptor mediated targeting, such as those of Avant Immunotherapeutics, Inc. (Needham, MA) may also be used.

Vaccine compositions often include adjuvants. Many adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, Bortadella pertussis or Mycobacterium tuberculosis derived proteins. Certain adjuvants are commercially available as, e.g., Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA);

aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A; and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Vaccines can be administered as nucleic acid compositions wherein DNA or RNA encoding one or more of the polypeptides, or a fragment thereof, is administered to a patient. See, e.g., Wolff, et al. (1990) Science 247:1465-1468 as well as US Patent Nos. 5,580,859; 5,589,466; 5,804,566; 5,739,118; 5,736,524; 5,679,647; WO 98/04720; and in more detail below. Examples of DNA-based delivery technologies include "naked DNA", facilitated (bupivicaine, polymers, peptide-mediated) delivery, cationic lipid complexes, and particle-mediated ("gene gun") or pressure-mediated delivery (see, e.g., US Patent No. 5,922,687).

For therapeutic or prophylactic immunization purposes, the peptides of the invention can be expressed by viral or bacterial vectors. Examples of expression vectors include attenuated viral hosts, such as vaccinia or fowlpox. This approach involves the use of vaccinia virus, e.g., as a vector to express nucleotide sequences that encode cancer polypeptides or polypeptide fragments. Upon introduction into a host, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits an immune response. Vaccinia vectors and methods useful in immunization protocols are described, e.g., in US Patent No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). See Stover, et al. (1991) Nature 351:456-460. A wide variety of other vectors are available for therapeutic administration or immunization, e.g., adeno and adeno-associated virus vectors, retroviral vectors, Salmonella typhi vectors, detoxified anthrax toxin vectors, and the like. See, e.g., Shata, et al. (2000) Mol. Med. Today 6:66-71; Shedlock, et al. (2000) J. Leukoc. Biol. 68:793-806; and Hipp, et al. (2000) In Vivo 14:571-85.

Methods for the use of genes as DNA vaccines are well known, and include placing a cancer gene or portion of a cancer gene under the control of a regulatable promoter or a tissue-specific promoter for expression in a soft tissue sarcoma cancer patient. The soft tissue sarcoma cancer gene used for DNA vaccines can encode full-length soft tissue sarcoma cancer proteins, but more preferably encodes portions of the cancer proteins including peptides derived from the cancer protein. In one

embodiment, a patient is immunized with a DNA vaccine comprising a plurality of nucleotide sequences derived from a soft tissue sarcoma cancer gene. For example, soft tissue sarcoma cancer-associated genes or sequence encoding subfragments of a soft tissue sarcoma cancer protein are introduced into expression vectors and tested for their immunogenicity in the context of Class I MHC and an ability to generate cytotoxic T cell responses. This procedure provides for production of cytotoxic T cell responses against cells which present antigen, including intracellular epitopes.

In one embodiment, the DNA vaccines include a gene encoding an adjuvant molecule with the DNA vaccine. Such adjuvant molecules include cytokines that increase the immunogenic response to the soft tissue sarcoma cancer polypeptide encoded by the DNA vaccine. Additional or alternative adjuvants are available.

In another embodiment soft tissue sarcoma cancer genes find use in generating animal models of soft tissue sarcoma cancer. When the soft tissue sarcoma cancer gene identified is repressed or diminished in cancer tissue, gene therapy technology, e.g., wherein antisense RNA directed to the soft tissue sarcoma cancer gene will also diminish or repress expression of the gene. Animal models of soft tissue sarcoma cancer find use in screening for modulators of a soft tissue sarcoma cancer-associated sequence or modulators of soft tissue sarcoma cancer. Similarly, transgenic animal technology including gene knockout technology, e.g., as a result of homologous recombination with an appropriate gene targeting vector, will result in the absence or increased expression of the soft tissue sarcoma cancer protein. When desired, tissue-specific expression or knockout of the soft tissue sarcoma cancer protein may be necessary.

It is also possible that the soft tissue sarcoma cancer protein is overexpressed in soft tissue sarcoma cancer. As such, transgenic animals can be generated that overexpress the soft tissue sarcoma cancer protein. Depending on the desired expression level, promoters of various strengths can be employed to express the transgene. Also, the number of copies of the integrated transgene can be determined and compared for a determination of the expression level of the transgene. Animals generated by such methods find use as animal models of soft tissue sarcoma cancer and are additionally useful in screening for modulators to treat soft tissue sarcoma cancer or to evaluate a therapeutic entity.

Kits for Use in Diagnostic and/or Prognostic Applications

For use in diagnostic, research, and therapeutic applications suggested above, kits are also provided by the invention. In the diagnostic and research applications such kits may include at least one of the following: assay reagents, buffers, soft tissue sarcoma cancer-specific nucleic acids or antibodies, hybridization probes and/or primers, antisense polynucleotides, ribozymes, dominant negative soft tissue sarcoma cancer polypeptides or polynucleotides, small molecules inhibitors of cancer-associated sequences, etc. A therapeutic product may include sterile saline or another pharmaceutically acceptable emulsion and suspension base.

In addition, the kits may include instructional materials containing directions (e.g., protocols) for the practice of the methods of this invention. While the instructional materials typically comprise written or printed materials they are not limited to such. A medium capable of storing such instructions and communicating them to an end user is contemplated by this invention. Such media include, but are not limited to electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. Such media may include addresses to internet sites that provide such instructional materials.

The present invention also provides for kits for screening for modulators of soft tissue sarcoma cancer-associated sequences. Such kits can be prepared from readily available materials and reagents. For example, such kits can comprise one or more of the following materials: a soft tissue sarcoma cancer-associated polypeptide or polynucleotide, reaction tubes, and instructions for testing cancer-associated activity. Optionally, the kit contains biologically active soft tissue sarcoma cancer protein. A wide variety of kits and components can be prepared according to the present invention, depending upon the intended user of the kit and the particular needs of the user. Diagnosis would typically involve evaluation of a plurality of genes or products. The genes will be selected based on correlations with important parameters in disease which may be identified in historical or outcome data.

EXAMPLES

Example 1

Tissue Preparation, Labeling Chips, and Fingerprints

Purify total RNA from tissue using TRIZOL® Reagent

Estimate tissue weight. Homogenize tissue samples in 1ml of TRIZOL® (chemical reagents for use in isolating biological material from organic tissue) per 50mg of tissue using a Polytron 3100 homogenizer. The generator/probe used depends upon the tissue size. A generator that is too large for the amount of tissue to be homogenized will cause a loss of sample and lower RNA yield. Use the 20mm generator for tissue weighing more than 0.6g. If the working volume is greater than 2ml, then homogenize tissue in a 15ml polypropylene tube (Falcon 2059). Fill tube no greater than 10ml.

HOMOGENIZATION

Before using generator, it should have been cleaned after last usage by running it through soapy H20 and rinsing thoroughly. Run through with EtOH to sterilize. Keep tissue frozen until ready. Add TRIZOL® directly to frozen tissue then homogenize.

Following homogenization, remove insoluble material from the homogenate by centrifugation at 7500 x g for 15 min. in a Sorvall superspeed or 12,000 x g for 10 min. in an Eppendorf centrifuge at 4°C. Transfer the cleared homogenate to a new tube(s). The samples may be frozen now at -60 to -70°C (and kept for at least one month) or you may continue with the purification.

PHASE SEPARATION

Incubate the homogenized samples for 5 minutes at room temperature. Add 0.2ml of chloroform per 1ml of TRIZOL® reagent used in the original homogenization.

Cap tubes securely and shake tubes vigorously by hand (do not vortex) for 15 seconds.

Incubate samples at room temp. for 2-3 minutes. Centrifuge samples at 6500rpm in a Sorvall superspeed for 30 min. at 4°C. (You may spin at up to 12,000 x g for 10 min. but you risk breaking your tubes in the centrifuge.)

RNA PRECIPITATION

Transfer the aqueous phase to a fresh tube. Save the organic phase if isolation of DNA or protein is desired. Add 0.5ml of isopropyl alcohol per 1ml of TRIZOL® reagent used in the original homogenization. Cap tubes securely and invert to mix.

Incubate samples at room temp. for 10 minutes. Centrifuge samples at 6500rpm in Sorvall for 20min. at 4°C.

RNA WASH

Pour off the supernate. Wash pellet with cold 75% ethanol. Use 1ml of 75% ethanol per 1ml of TRIzol reagent used in the initial homogenization. Cap tubes securely and invert several times to loosen pellet. (Do not vortex). Centrifuge at <8000rpm (<7500 x g) for 5 minutes at 4°C.

Pour off the wash. Carefully transfer pellet to an eppendorf tube (let it slide down the tube into the new tube and use a pipet tip to help guide it in if necessary). Depending on the volumes you are working with, you can decide what size tube(s) you want to precipitate the RNA in. When I tried leaving the RNA in the large 15ml tube, it took so long to dry (i.e. it did not dry) that I eventually had to transfer it to a smaller tube. Let pellet dry in hood. Resuspend RNA in an appropriate volume of DEPC H_20. Try for 2-5ug/ul. Take absorbance readings.

Purify poly A+ mRNA from total RNA or clean up total RNA with Qiagen's RNeasy kit

Purification of poly A^+ mRNA from total RNA. Heat oligotex suspension to 37°C and mix immediately before adding to RNA. Incubate Elution Buffer at 70°C. Warm up 2 x Binding Buffer at 65°C if there is precipitate in the buffer. Mix total RNA with DEPC-treated water, 2 x Binding Buffer, and Oligotex according to Table 2 on page 16 of the Oligotex Handbook. Incubate for 3 minutes at 65°C. Incubate for 10 minutes at room temperature.

Centrifuge for 2 minutes at 14,000 to 18,000 g. If centrifuge has a "soft setting," then use it. Remove supernatant without disturbing Oligotex pellet. A little bit of solution can be left behind to reduce the loss of Oligotex. Save sup until certain that satisfactory binding and elution of poly A^+ mRNA has occurred.

Gently resuspend in Wash Buffer OW2 and pipet onto spin column. Centrifuge the spin column at full speed (soft setting if possible) for 1 minute.

Transfer spin column to a new collection tube and gently resuspend in Wash Buffer OW2 and centrifuge as describe herein.

Transfer spin column to a new tube and elute with 20 to 100 ul of preheated (70°C) Elution Buffer. Gently resuspend Oligotex resin by pipetting up and down. Centrifuge as above. Repeat elution with fresh elution buffer or use first eluate to keep the elution volume low.

Read absorbance, using diluted Elution Buffer as the blank.

Before proceeding with cDNA synthesis, the mRNA must be precipitated. Some component leftover or in the Elution Buffer from the Oligotex purification procedure will inhibit downstream enzymatic reactions of the mRNA.

Ethanol Precipitation

Add 0.4 vol. of 7.5 M NH_4OAc + 2.5 vol. of cold 100% ethanol. Precipitate at -20°C 1 hour to overnight (or 20-30 min. at -70°C). Centrifuge at 14,000-16,000 x g for 30 minutes at 4°C. Wash pellet with 0.5ml of 80%ethanol (-20°C) then centrifuge at 14,000-16,000 x g for 5 minutes at room temperature. Repeat 80% ethanol wash. Dry the last bit of ethanol from the pellet in the hood. (Do not speed vacuum dry). Suspend pellet in DEPC H_20 at 1ug/ul concentration.

Clean up total RNA using Qiagen's RNeasy kit

Add no more than 100ug to an RNeasy column. Adjust sample to a volume of 100ul with RNase-free water. Add 350ul Buffer RLT then 250ul ethanol (100%) to the sample. Mix by pipetting (do not centrifuge) then apply sample to an RNeasy mini spin column. Centrifuge for 15 sec at >10,000rpm. If concerned about yield, re-apply flowthrough to column and centrifuge again.

Transfer column to a new 2-ml collection tube. Add 500ul Buffer RPE and centrifuge for 15 sec at >10,000rpm. Discard flowthrough. Add 500ul Buffer RPE and centrifuge for 15 sec at >10,000rpm. Discard flowthrough then centrifuge for 2 min at maximum speed to dry column membrane. Transfer column to a new 1.5-ml collection tube and apply 30-50ul of RNase-free water directly onto column membrane. Centrifuge 1 min at >10,000 rpm. Repeat elution.

Take absorbance reading. If necessary, ethanol precipitate with ammonium acetate and 2.5X volume 100% ethanol.

Make cDNA using Gibco's "SuperScript Choice System for cDNA Synthesis" kit

First Strand cDNA Synthesis

Use 5ug of total RNA or 1ug of polyA+ mRNA as starting material. For total RNA, use 2ul of SuperScript RT. For polyA+ mRNA, use 1ul of SuperScript RT. Final volume of first strand synthesis mix is 20ul. RNA must be in a volume no greater than 10ul. Incubate RNA with 1ul of 100pmol T7-T24 oligo for 10 min at 70C. On ice, add 7 ul of: 4ul 5X 1st Strand Buffer, 2ul of 0.1M DTT, and 1 ul of 10mM dNTP mix. Incubate at 37C for 2 min then add SuperScript RT Incubate at 37C for 1 hour.

Second Strand Synthesis

Place 1st strand reactions on ice.

Add: 91ul DEPC H20
30ul 5X 2nd Strand Buffer
3ul 10mM dNTP mix
1ul 10U/ul *E.coli* DNA Ligase
4ul 10U/ul *E.coli* DNA Polymerase
1ul 2U/ul RNase H

Make the above into a mix if there are more than 2 samples. Mix and incubate 2 hours at 16C.

Add 2ul T4 DNA Polymerase. Incubate 5 min at 16C. Add 10ul of 0.5M EDTA

Clean up cDNA

Phenol:Chloroform:Isoamyl Alcohol (25:24:1) purification using Phase-Lock gel (PLG) tubes.

Centrifuge PLG tubes for 30 sec at maximum speed. Transfer cDNA mix to PLG tube. Add equal volume of phenol:chloroform:isoamyl alcohol and shake vigorously (do not vortex). Centrifuge 5 minutes at maximum speed. Transfer top aqueous solution to a new tube. Ethanol precipitate: add 7.5X 5M NH4Oac and 2.5X volume of 100% ethanol. Centrifuge immediately at room temp. for 20 min, maximum speed. Remove sup then wash pellet 2X with cold 80% ethanol. Remove as much ethanol wash as possible then let pellet air dry. Resuspend pellet in 3ul RNase-free water.

In vitro Transcription (IVT) and labeling with biotin

Pipet 1.5ul of cDNA into a thin-wall PCR tube.

Make NTP labeling mix:

Combine at room temperature:	2ul	T7 10xATP (75mM) (Ambion)
	2ul	T7 10xGTP (75mM) (Ambion)
	1.5ul	T7 10xCTP (75mM) (Ambion)
	1.5ul	T7 10xUTP (75mM) (Ambion)
	3.75ul	10mM Bio-11-UTP (Boehringer-Mannheim/Roche or Enzo)
	3.75ul	10mM Bio-16-CTP (Enzo)

2ul 10x T7 transcription buffer (Ambion)

2ul 10x T7 enzyme mix (Ambion)

Final volume of total reaction is 20ul. Incubate 6 hours at 37C in a PCR machine.

RNeasy clean-up of IVT product

Follow previous instructions for RNeasy columns or refer to Qiagen's RNeasy protocol handbook.

cRNA will most likely need to be ethanol precipitated. Resuspend in a volume compatible with the fragmentation step.

Fragmentation

15 ug of labeled RNA is usually fragmented. Try to minimize the fragmentation reaction volume; a 10 ul volume is recommended but 20 ul is all right. Do not go higher than 20 ul because the magnesium in the fragmentation buffer contributes to precipitation in the hybridization buffer.

Fragment RNA by incubation at 94 C for 35 minutes in 1 x Fragmentation buffer.

5 x Fragmentation buffer:

200 mM Tris-acetate, pH 8.1

500 mM KOAc

150 mM MgOAc

The labeled RNA transcript can be analyzed before and after fragmentation. Samples can be heated to 65C for 15 minutes and electrophoresed on 1% agarose/TBE gels to get an approximate idea of the transcript size range

Hybridization

200 ul (10ug cRNA) of a hybridization mix is put on the chip. If multiple hybridizations are to be done (such as cycling through a 5 chip set), then it is recommended that an initial hybridization mix of 300 ul or more be made.

Hybrization Mix: fragment labeled RNA (50ng/ul final conc.)

50 pM 948-b control oligo

1.5 pM BioB

5 pM BioC

25 pM BioD

100 pM CRE

0.1mg/ml herring sperm DNA

0.5mg/ml acetylated BSA

to 300 ul with 1xMES hyb. buffer

The instruction manuals for the products used herein are incorporated herein in their entirety.

Labeling Protocol

Hybridization reaction:

Start with non-biotinylated IVT (purified by RNeasy columns)

(see example 1 for steps from tissue to IVT)

IVT antisense RNA; 4 μg:	μl
Random Hexamers (1 μg/μl):	4 μl
H_2O:	μl
	14 μl

- Incubate 70°C, 10 min. Put on ice.

Reverse transcription:

5X First Strand (BRL) buffer:	*6 μl*
0.1 M DTT:	3 μl
50X dNTP mix:	0.6 μl
H2O:	2.4 μl
Cy3 or Cy5 dUTP (1mM):	3 μl
SS RT II (BRL):	1 μl
	16 μl

- Add to hybridization reaction.
- Incubate 30 min., 42°C.
- Add 1 μl SSII and let go for another hour.

Put on ice.

- 50X dNTP mix (25mM of cold dATP, dCTP, and dGTP, 10mM of dTTP: 25 μl each of 100mM dATP, dCTP, and dGTP; 10 μl of 100mM dTTP to 15 μl H2O. dNTPs from Pharmacia)

RNA degradation:

	86 μl H_2O
- Add 1.5 μl 1M NaOH/ 2mM EDTA, incubate at 65°C, 10 min.	10 μl 10N NaOH

4 µl 50mM

EDTA

U-Con 30

500 µl TE/sample spin at 7000g for 10 min, save flow through for purification

Qiagen purification:

-suspend u-con recovered material in 500µl buffer PB

-proceed w/ normal Qiagen protocol

DNAse digest:

- Add 1 µl of 1/100 dil of DNAse/30µl Rx and incubate at 37°C for 15 min.

-5 min 95°C to denature enzyme

Sample preparation:

- Add:
- Cot-1 DNA: 10 µl

50X dNTPs: 1 µl

20X SSC: 2.3 µl

Na pyro phosphate: 7.5 µl

10mg/ml Herring sperm DNA 1ul of 1/10 dilution

21.8 final vol.

- Dry down in speed vac.
- Resuspend in 15 µl H_20.
- Add 0.38 µl 10% SDS.
- Heat 95°C, 2 min.
- Slow cool at room temp. for 20 min.

Put on slide and hybridize overnight at 64°C.

Washing after the hybridization:

3X SSC/0.03% SDS: 2 min. 37.5 mls 20X SSC+0.75mls 10% SDS in 250mls H_2O

1X SSC: 5 min. 12.5 mls 20X SSC in 250mls H_2O

0.2X SSC: 5 min. 2.5 mls 20X SSC in 250mls H_2O

Dry slides in centrifuge, 1000 RPM, 1min.

Scan at appropiate PMT's and channels.

Example 2: Gene Chip Analysis

Molecular profiles of various normal and soft tissue sarcoma cancer tissues were determined and analyzed using gene chips. RNA was isolated and gene chip analysis was performed as described above (Glynne, et al. (2000) Nature 403:672-676; Zhao, et al. (2000) Genes Dev. 14:981-993). The results are shown in the tables and figures that follow. These soft tissue sarcoma (STSD) associated sequences are identified in the tables by Genbank Accession numbers and gene titles. As indicated, some of the Accession numbers include expression sequence tags (ESTs). Thus, in one embodiment herein, genes within an expression profile, also termed expression profile genes, include ESTs and are not necessarily full length.

It is understood that the examples described above in no way serve to limit the true scope of this invention to specific embodiments, but rather are presented for illustrative purposes. All publications, sequences of accession numbers, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.